

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:56:42 ; Search time 22.16 Seconds  
(without alignments)  
582.982 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGPLPRTVELFYDLSPYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 390729 seqs, 57163235 residues

Word size : 0

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq\_0401.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	226	100.0	226	21	Y77499	Novel human glutat
2	226	100.0	256	20	Y59988	Human endometrium
3	131	58.0	131	21	G01523	Human secreted pro
4	97	42.9	97	21	B34372	Human secreted pro
5	88	38.9	300	21	B43298	Human ORFX ORF3062
6	12	5.3	97	21	B34971	Gene 45 human secr
7	8	3.5	179	21	B58191	Lung cancer associ
8	3	3.5	195	21	B12526	Human Ma2 protein
9	8	3.5	283	21	B12528	Human Ma4 protein
10	8	3.5	813	15	B62254	SefC sequence. Sa
11	8	3.5	814	18	W23573	Salmonella enterit

12	7	3.1	55	21	B09366	Hepatitis GB virus
13	7	3.1	63	17	R99624	P. aeruginosa oprl
14	7	3.1	83	11	R05708	Pseudomonas aerugi
15	7	3.1	130	21	B58516	Lung cancer associ
16	7	3.1	131	13	R22260	Profilin, P14(T) a
17	7	3.1	131	13	Z55643	Phleum sp. allerg
18	7	3.1	131	20	W78443	P.pratense pollen
19	7	3.1	131	20	W78443	P.pratense pollen
20	7	3.1	131	20	W78443	P14 allergen of bl
21	7	3.1	133	13	R22261	Birch pollen P14 a
22	7	3.1	133	16	R71926	Betula sp. allerg
23	7	3.1	133	20	W25650	Human pollen all
24	7	3.1	133	20	W78442	M.annua pollen all
25	7	3.1	133	20	W25770	Human hds3 ul. Ho
26	7	3.1	204	18	W54075	Human tumour prote
27	7	3.1	204	19	W54075	Human D53 protein.
28	7	3.1	204	20	Y29626	Mouse D53 protein.
29	7	3.1	204	21	Y49953	Mouse D53 protein.
30	7	3.1	214	17	R99626	P. aeruginosa orpf
31	7	3.1	226	17	R99627	Human glutamate re
32	7	3.1	292	16	R84918	Human glutamate re
33	7	3.1	292	16	R84919	Neisseria meningit
34	7	3.1	384	21	Y75203	Neisseria meningit
35	7	3.1	389	21	Y75202	Neisseria meningit
36	7	3.1	397	12	R14187	S. lipmanli epimer
37	7	3.1	650	15	R54089	Partial sequence o
38	7	3.1	733	21	B41232	Human ORFX ORF996
39	7	3.1	854	19	W79152	Receptor protein t
40	7	3.1	854	20	W81409	Receptor protein t
41	7	3.1	855	16	R75503	Human colonic aden
42	7	3.1	855	16	R75503	Human colonic aden
43	7	3.1	855	18	W34671	CCK-2, a human mam
44	7	3.1	855	19	W77114	Discoidin domain f
45	7	3.1	868	20	Y31726	Human fused CDNA 1

## ALIGNMENTS

RESULT 1  
Y77499  
ID Y77499 standard; Protein; 226 AA.  
XX  
AC Y77499;  
XX  
DT 05-JUN-2000 (first entry)  
XX  
DE Novel human glutathione S-transferase, GSTS.  
XX  
KW Glutathione S-transferase; human; GSTS; cancer; immune disorder;  
KW gene therapy; diagnosis; treatment; drug screening.  
XX  
OS Homo sapiens.  
XX  
PN US6030809-A.  
XX  
PD 29-FEB-2000.  
XX  
PF 25-NOV-1997; 97US-0978174.  
XX  
PR 25-NOV-1997; 97US-0978174.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Shah P, Lal P, Corley NC;  
XX  
DR WPI: 2000-205204/18.  
XX  
DR N-PSDB; 202599.  
XX  
PT Isolated nucleic acid encoding glutathione S-transferase useful in the  
PT production of agents for preventing, diagnosing and treating diseases  
PT associated with cell proliferation -  
XX

PS Claim 8; Fig 1A-C; 27pp; English.

XX This sequence represents a novel human glutathione S-transferase, GSTS.  
 CC Nucleotide sequences encoding GSTS were initially isolated from a  
 CC urologic cDNA library, and subsequently extended using cDNA libraries  
 CC derived from other tissues, such as brain or bladder. The present  
 CC sequence is encoded by a consensus cDNA. GSTS, and nucleotides which  
 CC encode it may be used in the prevention, treatment and diagnosis of  
 CC diseases associated with inappropriate GSTS expression, such as cancers  
 CC and immune disorders. Nucleotides which encode GSTS may be used in gene  
 CC therapy to treat disorders associated with reduced expression or activity  
 CC of GSTS, and in antisense therapy for disorders associated with increased  
 CC GSTS expression or activity. They may also be used for the recombinant  
 CC production of GSTS, and as a source of probes and primers to detect and  
 CC quantitate the presence of similar nucleic acid sequences, particularly  
 CC for the diagnosis of GSTS-associated disorders. GSTS proteins may be used  
 CC as antigens in the production of antibodies against GSTS and in assays to  
 CC identify modulators (agonists and antagonists) of GSTS expression and  
 CC activity. The anti-GSTs antibodies and GSTS antagonists may also be used  
 CC to downregulate GSTS expression and activity. Antagonists of GSTS  
 CC expression and function may be used to treat immune disorders (e.g., AIDS,  
 CC anaemia, asthma, Crohn's disease, irritable bowel syndrome, multiple  
 CC sclerosis, osteoarthritis, microbial infections) and cancers (e.g.,  
 CC leukaemia, lymphoma, melanoma, and cancers of the breast, prostate,  
 CC liver, lung and brain). The anti-GSTs antibodies may also be used as  
 CC diagnostic agents.

XX SQ Sequence 226 AA;

Query Match 100.0%; Score 226; DB 21; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-219;  
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSVSWLGFELICRYQNIWNINQLRPSLTIGIMKDSGNKPPGLLP 60  
 DB 1 mgplprtvelfdvlspsvswlgsfelicryqniwninqlrpsltigimkdsngkppgllp 60  
 QY 61 RKGLYMANDLKLRLHLLQIPIHPKDFLSVMEKSGLSAMRFLTAVNLEHPMELEKASRE 120  
 DB 61 rkglymandlklrlhllqihpfpkdfslsvmeksglsamrfltavnlhepemelekasre 120  
 QY 121 LWMRVWSRNEDETEPQSILAAAEKAGMSAEQAGLLEKIATPKVNOLKETTEAACRYGA 180  
 DB 121 lwmrwvwsrneditepqsilaaekagmsaeqaglllekiatpkvknolketteaacryga 180  
 QY 181 FGLPITVAHVDPGQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226  
 DB 181 fgipitvahvdgqthmlfgsdrnellahllgkwmgpippavnarl 226

RESULT 2  
 Y59988  
 ID Y59988 standard; Protein: 256 AA.

XX Y59988;

XX 31-JAN-2000 (first entry)

DE Human endometrium tumour EST encoded protein 48.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
 KW treatment; uterine; gene therapy; expressed sequence tag.

XX Homo sapiens.

XX DE19817948-A1.

XX 21-OCT-1999.

XX 17-APR-1998; 98DE-1017948.

XX 17-APR-1998; 98DE-1017948.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
 XX WPI; 1999-591957/31.  
 XX N-PSDB; 241996.

XX New nucleic acid sequences expressed in uterine cancer tissues, and  
 XX derived polypeptides, for treatment of uterine and endometrial cancer  
 XX and identification of therapeutic agents -  
 XX Claim 23; Page 294; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for generation  
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
 CC sequence tags) from a particular tissue type before comparison of  
 CC expression patterns. This allows a significantly longer fragment of the  
 CC gene to be revealed, so should reduce the number of failures associated  
 CC with the fact that ESTs from different libraries may represent different  
 CC parts of the same unknown gene, distorting the estimated frequency of  
 CC occurrence in a particular tissue. Y59941-Y60328 represent protein  
 CC fragments encoded by the human endometrium tumour cDNA library derived  
 CC EST fragments represented in Z41981-Z42121.

XX SQ Sequence 256 AA;

Query Match 100.0%; Score 226; DB 20; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-219;  
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSVSWLGFELICRYQNIWNINQLRPSLTIGIMKDSGNKPPGLLP 60  
 DB 31 mgplprtvelfdvlspsvswlgsfelicryqniwninqlrpsltigimkdsngkppgllp 90  
 QY 61 RKGLYMANDLKLRLHLLQIPIHPKDFLSVMEKSGLSAMRFLTAVNLEHPMELEKASRE 120  
 DB 61 rkglymandlklrlhllqihpfpkdfslsvmeksglsamrfltavnlhepemelekasre 150  
 QY 121 LWMRVWSRNEDETEPQSILAAAEKAGMSAEQAGLLEKIATPKVNOLKETTEAACRYGA 180  
 DB 151 lwmrwvwsrneditepqsilaaekagmsaeqaglllekiatpkvknolketteaacryga 210  
 QY 181 FGLPITVAHVDPGQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226  
 DB 211 fgipitvahvdgqthmlfgsdrnellahllgkwmgpippavnarl 256

RESULT 3  
 G01523

ID G01523 standard; Protein: 131 AA.

XX G01523;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 5604.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.  
 XX 26-FEB-1999; 99US-0122487.  
 XX (GEST ) GENSET.  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX WPI: 2000-500381/45.  
 XX N-PSDB; C01529.  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 13; SEQ ID 5604; 71pp + CD-ROM; English.  
 XX The present sequence is a polypeptide encoded by one of a large number  
 XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 XX were prepared from total human RNAs or polyA+ RNAs derived from 30  
 XX different tissues. EST sequences usually correspond mainly to the 3'  
 XX untranslated region (UTR) of the mRNA because they are often obtained  
 XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 XX those cases where longer cDNA sequences have been obtained, the full 5'  
 XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 XX ends and can therefore be used to obtain full length cDNAs and genomic  
 XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 XX chromosome mapping procedures. They are used to obtain upstream  
 XX regulatory sequences and to design expression and secretion vectors.  
 XX Sequence 131 AA;  
 XX  
 Query Match 58.0%; Score 131; DB 21; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-124;  
 Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPLPRTVELFDVLSYSGWIGFELCRQYQNIWNLRLPSLITGIMKDSGNKPPGLP 60  
 DB 1 mgplprtvelfdvlsysgwifellcrqynlwnlnlgrpslitgimkdsngnkppllp 60  
 QY 61 RKGLYMANDKLLRHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLHPMLEKASRE 120  
 DB 61 rkglymandkllrhhlqipfhfpkdfslsvmleksglsamrfltavnlhpemlekasre 120  
 QY 121 LWMRVWSRNE 131  
 DB 121 lwmrwvrsrned 131  
 RESULT 4  
 B34972  
 ID B34972 standard; Protein; 97 AA.  
 XX AC B34972;  
 XX 26-JAN-2001 (first entry)  
 XX Human secreted protein sequence encoded by gene 45 SEQ ID NO:176.  
 XX Human: secreted protein; neuroprotective; cytosolic; cardioactive;  
 XX immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;  
 XX antiinfective; gynaecological; antibacterial; neural disorder; cancer;  
 XX immune disease; reproductive disorder; proliferative disorder;  
 XX gastrointestinal disease; wound healing; infectious disease;  
 XX food additive.  
 XX Homo sapiens.  
 XX WO200056766-A1.  
 XX

PD 28-SEP-2000.  
 XX 16-MAR-2000; 2000WO-US06824.  
 XX 19-MAR-1999; 99US-0125359.  
 XX 03-DEC-1999; 99US-0168664.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM, Komatsoulis G;  
 XX WPI: 2000-594574/56.  
 XX N-PSDB; C60069.  
 XX Human secreted proteins and gene sequences encoding them, useful for  
 XX detection, prevention, and treatment of various disorders such as  
 XX cancer and immune system disorders -  
 XX Disclosure; Page 83; 442pp; English.  
 XX The polynucleotide sequences given in C60025-C60071 encode the human  
 XX secreted proteins represented in B34854-B34900. Sequences B34901-B34976  
 XX are fragments of proteins encoded by the genes, and also proteins which  
 XX which they share sequence homology. The proteins have activities based on  
 XX the tissues in which their encoding genes are expressed. Examples of  
 XX the proteins activities include: neuroprotective; cytosolic;  
 XX cardioactive; immunomodulatory; general muscular activity; vulnary;  
 XX general gastrointestinal activity; nephrotropic; antiinfective;  
 XX gynaecological; and antibacterial. The human secreted proteins,  
 XX polynucleotides, antagonists and agonists of the invention may be  
 XX useful in treating, preventing and/or diagnosing various diseases,  
 XX disorders and conditions such as neural, immune, muscular, reproductive,  
 XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 XX disorders and cancer. They may also be used in the treatment of wounds,  
 XX and infectious diseases. The polypeptides may be used as a food additive  
 XX or preservative to increase storage capabilities. Sequences C60016-C60024  
 XX and B34853 are used in the course of the invention during the  
 XX identification and characterisation of the protein and nucleotide  
 XX sequences.  
 XX Sequence 97 AA;  
 XX  
 Query Match 42.9%; Score 97; DB 21; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-90;  
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 130 EDITPEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNOLKETTAAACRYGAFGLPITVAH 189  
 DB 1 editpepqsilaaekagmsaeqglllekiatpkvknqlketteaacrygafglpityah 60  
 QY 190 VDGQTHMLFGSDRMELLALLGKWKWGPPIPPAVNARL 226  
 DB 61 vdgqthmlfgsdrnellallgkwmgpippavnarl 97  
 RESULT 5  
 B43298  
 ID B43298 standard; Protein; 300 AA.  
 XX AC B43298;  
 XX 08-FEB-2001 (first entry)  
 XX Human ORFX ORF3062 polypeptide sequence SEQ ID NO:6124.  
 XX Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;  
 XX vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 XX immunostimulant; thrombolytic; coagulant; vasotrophic; antiinflammatory;  
 XX hypotensive; dermatological; immunosuppressive; antitumor; antitubercular;  
 XX antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;  
 XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 XX

neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.  
WO200058473-A2.  
05-OCT-2000.  
31-MAR-2000; 2000WO-US08621.  
31-MAR-1999; 99US-0127607.  
02-APR-1999; 99US-0127636.  
05-APR-1999; 99US-0127728.  
30-MAR-2000; 2000US-0540763.  
(CURA-) CURAGEN CORP.  
Shimkets RA, Leach M;  
WPI: 2000-602362/57.  
N-PSDB; C77507.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -  
Claim 11; Page 5304-5305; 5507pp; English.

C74446 to C77606 encode the proteins given in B40237 to B43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antineoplastic; antibacterial; antiviral; antifungal; antihemagic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 300 AA;

Query Match 38.9%; Score 88; DB 21; Length 300;  
Best Local Similarity 100.0%; Pred. No. 2.1e-80;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 NEDITEPQSTILAAAEKAGMSAQOGLLEKIKATPKVKNOLKETTEACRYGALPLITVA 188  
|||||  
Db 213 neditepqslaaekagmsaqoaglllekikatkpkvknolketteaacrygalfpitva 272  
QY 189 HVDGQTHMFGSDRMELLALHLLGKWMG 216  
|||||  
Db 273 hvdgqthmlfgsdrmlallllgkwmg 300

RESULT 6  
B34971

ID B34971 standard; Protein; 97 AA.  
XX B34971;  
AC  
XX  
DT 26-JAN-2001 (first entry)  
XX  
DE Gene 45 human secreted protein homologous amino acid sequence #175.  
XX  
XX Human; secreted protein; neuroprotective; cytostatic; cardioactive;  
KW immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;  
KW antiinfective; gynaecological; antibacterial; neural disorder; cancer;  
KW immune disease; reproductive disorder; proliferative disorder;  
KW gastrointestinal disease; wound healing; infectious disease;  
KW food additive.  
XX  
OS Rattus sp.  
XX  
XX WO200056766-A1.  
PN  
XX  
XX 28-SEP-2000.  
PD  
XX  
XX 16-MAR-2000; 2000WO-US06824.  
XX  
XX 19-MAR-1999; 99US-0125359.  
PR  
XX 03-DEC-1999; 99US-0168664.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Komatsoulis G;  
PI  
XX WPI: 2000-594574/56.  
XX

Human secreted proteins and gene sequences encoding them, useful for detection, prevention, and treatment of various disorders such as cancer and immune system disorders -  
Disclosure; Page 83; 442pp; English.

The polynucleotide sequences given in C60025-C60071 encode the human secreted proteins represented in B34854-B34900. Sequences B34901-B34976 are fragments of proteins encoded by the genes, and also proteins with which they share sequence homology. The proteins have activities based on the tissues in which their encoding genes are expressed. Examples of the proteins activities include: neuroprotective; cytostatic; cardioactive; immunomodulatory; general muscular activity; vulnary; general gastrointestinal activity; general muscular activity; infective; gynaecological; and antibacterial. The human secreted proteins, polynucleotides, and antagonists and/or diagnosing various diseases, useful in treating, preventing and/or diagnosing various diseases, disorders and conditions such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders and cancer. They may also be used in the treatment of wounds, and infectious diseases. The polypeptides may be used as a food additive or preservative to increase storage capabilities. Sequences C60016-C60024 and B34853 are used in the course of the invention during the identification and characterisation of the protein and nucleotide sequences.

Sequence 97 AA;

Query Match 5.3%; Score 12; DB 21; Length 97;  
Best Local Similarity 100.0%; Pred. No. 0.00021;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 MLFGSDRMELLA 207  
|||||  
Db 67 mlfgsdrmlella 78

RESULT 7  
B58191  
ID B58191 standard; Protein; 179 AA.

DT 02-NOV-2000 (first entry)  
XX Human Ma2 protein SEQ ID NO:7.  
XX  
XX  
KW Ma1; Ma2; Ma3; Ma4; Ma5; paraneoplastic limbic encephalitis; diagnosis;  
KW paraneoplastic syndrome; neoplasm; colon cancer; brainstem encephalitis;  
KW breast cancer; parotid gland cancer; lung cancer; testicular cancer;  
KW germ-cell tumour.  
XX  
OS Homo sapiens.  
XX  
PN JP2000146982-A.  
XX  
PD 26-MAY-2000.  
XX  
PF 10-NOV-1999; 99JP-0320171.  
XX  
PR 10-NOV-1998; 98US-0189527.  
XX  
PA (SLOAN ) SLOAN KETTERING INST CANCER RES.  
XX  
DR WPI: 2000-468119/41.  
DR N-PSDB; A60834.  
XX  
PT Diagnosing paraneoplastic syndrome e.g. paraneoplastic limbic  
PT encephalitis or neoplasm e.g. colon cancer comprising assessing a test  
PT sample for the presence or absence of antibodies to a Ma family  
PT polypeptide  
XX  
PS Claim 48; Fig 2; 27pp; Japanese.  
XX  
CC The present invention describes a method for diagnosing a paraneoplastic  
CC syndrome or neoplasm. The method comprises assessing a test sample for  
CC the presence or absence of antibodies to a Ma family polypeptide (I).  
CC the method is used to diagnose a paraneoplastic syndrome especially  
CC paraneoplastic limbic encephalitis and/or brainstem encephalitis or  
CC neoplasm in an individual. The method diagnoses the neoplasm by assessing  
CC antibodies to (I) preferably Ma1, which is indicative presence of breast  
CC cancer, colon cancer, parotid gland cancer, lung cancer, testicular  
CC cancer and germ-cell tumours or Ma2, which is indicative of testicular  
CC cancer, germ-cell tumour, and lung cancer. The present sequence is  
CC the Ma2 protein as given in the present invention.  
XX  
SQ Sequence 195 AA;  
Query Match 3.5%; Score 8; DB 21; Length 195;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 204 ELLAHLG 211  
Db 139 ellahllg 146  
RESULT 9  
B12528  
ID B12528 standard; Protein; 283 AA.  
XX  
AC B12528;  
XX  
DT 02-NOV-2000 (first entry)  
XX  
DE Human Ma4 protein SEQ ID NO:11.  
XX  
KW Ma1; Ma2; Ma3; Ma4; Ma5; paraneoplastic limbic encephalitis; diagnosis;  
KW paraneoplastic syndrome; neoplasm; colon cancer; brainstem encephalitis;  
KW breast cancer; parotid gland cancer; lung cancer; testicular cancer;  
KW germ-cell tumour.  
XX  
OS Homo sapiens.  
XX  
PN JP2000146982-A.

XX B58191;  
AC  
XX  
DT 14-MAR-2001 (first entry)  
DE  
XX  
XX Lung cancer associated polypeptide sequence SEQ ID 529.  
XX  
XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
KW cardioactive; immunomodulatory; muscular active; vulnerary;  
KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200055180-A2.  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 08-MAR-2000; 2000WO-US05918.  
PF  
XX  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX (ROSE/) ROSEN C A.  
PA  
XX Ruben SM;  
PI  
XX  
XX WPI: 2000-587514/55.  
DR  
DR N-PSDB; F18067.  
XX  
XX Lung cancer associated gene sequences, referred to as lung cancer  
PT antigens, useful for treatment, prevention, and diagnosis of disorders  
PT such as lung cancer.  
XX  
XX Claim 11; Page 1021; 1425pp; English.  
PS  
XX Polynucleotide sequences F17982 - F18424 encode human lung cancer  
XX associated proteins represented in B58106 - B58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cytostatic; cardioactive;  
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
CC general; nephrotropic; antiinfective; gynecological; or antibacterial  
CC activity. The invention also includes antibodies specific for the protein  
CC or polynucleotide sequences. The lung cancer associated polynucleotide  
CC sequences may be used for detection of lung cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The proteins may be used to treat disorders such as  
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,  
CC cardiovascular, renal, and proliferative disorders. The proteins may also  
CC be used in the treatment of wounds and infectious diseases.  
CC Polynucleotide sequences F18425 - F18433 and peptide B58549 are used in  
CC the course of the invention for the identification and characterisation  
CC of the polynucleotide and protein sequences.  
XX  
XX Sequence 179 AA;  
Query Match 3.5%; Score 8; DB 21; Length 179;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 96 SLSAMRFL 103  
Db 25 slsamrfl 32  
RESULT 8  
B12526  
ID B12526 standard; Protein; 195 AA.  
XX  
XX B12526;  
AC  
XX

XX 26-MAY-2000.  
 PD XX  
 PF 10-NOV-1999; 99JP-0320171.  
 XX  
 PR 10-NOV-1998; 98US-0189527.  
 XX  
 PA (SLOAN ) SLOAN KETTERING INST CANCER RES.  
 XX  
 DR WPI; 2000-468119/41.  
 DR N-PSDB; A60836.  
 XX  
 XX Diagnosing paraneoplastic syndrome e.g. paraneoplastic limbic  
 PT encephalitis or neoplasm e.g. colon cancer comprising assessing a test  
 PT sample for the presence or absence of antibodies to a Ma family  
 PT polypeptide -  
 XX  
 PS Claim 48; Fig 7-8; 27pp; Japanese.  
 XX  
 CC The present invention describes a method for diagnosing a paraneoplastic  
 CC syndrome or neoplasm. The method comprises assessing a test sample for  
 CC the presence or absence of antibodies to a Ma family polypeptide (1).  
 CC The method is used to diagnose a paraneoplastic syndrome especially  
 CC paraneoplastic limbic encephalitis and/or brainstem encephalitis or  
 CC neoplasm in an individual. The method diagnoses the neoplasm by assessing  
 CC antibodies to (1) preferably Ma1, which is indicative presence of breast  
 CC cancer, colon cancer, parotid gland cancer, lung cancer, testicular  
 CC cancer and germ-cell tumours or Ma2, which is indicative of testicular  
 CC cancer, germ-cell tumour, and lung cancer. The present sequence is  
 CC the Ma4 protein as given in the present invention.  
 XX  
 SQ Sequence 283 AA;

Query Match 3.5%; Score 8; DB 21; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ELLAHLG 211  
 Db 59 ellahlhg 66  
 |||||  
 |||||

RESULT 10  
 R62754  
 ID R62754 standard; Protein; 813 AA.  
 XX  
 AC R62754;  
 XX  
 DT 26-JUN-1995 (first entry)  
 XX  
 DE SefC sequence.  
 XX  
 KW Salmonella; SefC; vaccine.  
 XX  
 OS Salmonella.  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 497  
 FT /note= "illegible in patent specification"  
 XX  
 PN W09425598-A.  
 XX  
 PD 10-NOV-1994.  
 XX  
 PF 26-APR-1994; 94WO-IB00207.  
 XX  
 PR 26-APR-1993; 93US-0054452.  
 XX  
 PA (KING/) KING J.  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 PI Clouthier SC, Collinson SK, Doran JL, Kay WW;

XX WPI; 1994-358275/44.  
 DR N-PSDB; Q73061.  
 XX  
 PT Eliciting an immune response to Salmonella - using attenuated  
 PT Salmonella strains, vector constructs, or compsns. contg.  
 PT fimbrial type proteins.  
 XX  
 PS Disclosure; Figure 2A-2D; 95pp; English.  
 XX  
 CC The isolated SefC protein may be used in a vaccine composition to  
 CC elicit an immune response to Salmonella in animals (e.g. food  
 CC producing animals) and humans.  
 XX  
 SQ Sequence 813 AA;

Query Match 3.5%; Score 8; DB 15; Length 813;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 NOLKETTE 173  
 Db 60 nqlkette 67  
 |||||  
 |||||

RESULT 11  
 W23573  
 ID W23573 standard; Protein; 814 AA.  
 XX  
 AC W23573;  
 XX  
 DT 29-SEP-1997 (first entry)  
 XX  
 DE Salmonella enteritidis sefC.  
 XX  
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 PN US5635617-A.  
 XX  
 PD 03-JUN-1997.  
 XX  
 PF 26-APR-1993; 93US-0054452.  
 XX  
 PR 26-APR-1994; 94US-0233788.  
 PR 26-APR-1993; 93US-0054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 PI Clouthier SC, Collinson SK, Doran JL, Kay WW;  
 XX  
 DR WPI; 1997-309886/28.  
 DR N-PSDB; T74143.  
 XX  
 PT Isolated Salmonella gene agfA - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteria family  
 XX  
 PS Example 9; Fig 2; 85pp; English.  
 XX  
 CC The present sequence represents sefC found in the sef gene cluster  
 CC from Salmonella enteritidis. The nucleic acid can be used to provide  
 CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of  
 CC the family Enterobacteria. It can also be used to provide proteins and  
 CC antibodies which can be used for assays. The nucleic acid sequence can  
 CC be used to provide probes or primers which can specifically hybridise  
 CC to nucleic acid molecules from greater than 99% of Salmonella strains  
 CC that are pathogenic to warm-blooded animals relative to nucleic acid  
 CC molecules from virtually all other microbial organisms.  
 XX  
 SQ Sequence 814 AA;

Query Match 3.1%; Score 8; DB 18; Length 814;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 NOLKETTE 173  
Db 60 nqlkette 67

RESULT 12  
B09366  
ID B09366 standard; Protein: 55 AA.  
XX B09366;  
XX  
DT 30-AUG-2000 (first entry)  
XX  
DE Hepatitis GB virus protein sequence SEQ ID NO:493.  
XX  
KW Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic;  
KW infection; detection; characterisation; hepatitis.  
XX  
OS Hepatitis GB virus.  
XX  
PN US6051374-A.  
XX  
PD 18-APR-2000.  
XX  
PF 07-JUN-1995; 95US-0488445.  
XX  
PR 14-FEB-1994; 94US-0196030.  
PR 13-MAY-1994; 94US-0242654.  
PR 29-JUL-1994; 94US-0283314.  
PR 23-NOV-1994; 94US-0344185.  
PR 23-NOV-1994; 94US-0344190.  
PR 30-JAN-1995; 95US-0377557.  
XX  
XX (ABBO ) ABBOTT LAB.  
XX  
PA Dawson GJ, Leary TP, Muerhoff AS, Pilot-Matias TJ, Buijk SL;  
PI Mushahwar IK, Simons JN, Desai SM, Erker JC, Schlauder GG;  
XX  
XX WPI; 2000-338307/29.  
XX  
PT Detecting target hepatitis GB virus nucleic acid in a test sample  
PT suspected of containing HGBV comprises reacting the test sample the  
PT HGBV polynucleotide probe and detecting the complex that contains  
PT target HGBV -  
XX  
PS Example 18; Column 521-522; 369pp; English.

XX  
XX The present invention describe a method for detecting target  
CC hepatitis GB virus (HGBV) nucleic acid (THN) in a test sample (T)  
CC suspected of containing HGBV. The method involves reacting (T) with a  
CC HGBV polynucleotide probe (I) containing 15 contiguous nucleotides, and  
CC which selectively hybridises to the HGBV genome or its full complement,  
CC and detecting the complex that contains THN, indicating the presence of  
CC target HGBV. The method is used for detecting target HGBV nucleic acid  
CC in the test sample suspected of containing HGBV and for characterisation  
CC of newly ascertained etiological agent of non-A, non-B, non-C, non-D and  
CC non-E hepatitis causing agents collectively termed as hepatitis GB virus.  
CC A55270 to A55489 and B08985 to B09480 represent nucleotide and protein  
CC sequences used in the exemplification of the present invention.  
XX  
XX Sequence 55 AA;  
SQ

Query Match 3.1%; Score 7; DB 21; Length 55;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 GSLSAMR 101

Db 42 gslsamr 48  
RESULT 13  
R99624  
ID R99624 standard; Protein: 63 AA.  
XX  
AC R99624;  
XX  
DT 01-OCT-1996 (first entry)  
XX  
DE P. aeruginosa oprI (without signal peptide).  
XX  
KW Vaccine; fusion protein; OprF; OprI; outer membrane protein;  
KW antibody; glutathione-S-transferase; GST.  
XX  
OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).  
XX  
PN EP717106-A1.  
XX  
PD 19-JUN-1996.  
XX  
PF 17-NOV-1995; 95EP-0118098.  
XX  
PR 16-DEC-1994; 94EP-0120023.  
XX  
PA (BEHW ) BEHRINGWERKE AG.  
XX  
PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;  
PI Von Specht BU;  
XX  
XX WPI; 1996-279559/29.  
DR N-PSDB; T32598.  
XX  
PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.  
PT fragments of outer membrane proteins I and F.  
XX  
PS Disclosure; Page 11; 23pp; English.

XX  
XX A cDNA sequence (T32598) codes for amino acids 21-83 (R99624) of  
CC the Pseudomonas aeruginosa ATCC 33354 outer membrane protein I  
CC (oprI), i.e. oprI lacking the signal peptide. A hybrid protein  
CC between glutathione-S-transferase and the OprI moiety was  
CC expressed in Escherichia coli. OprF-OprI (R99626) and OprI-OprF  
CC (R99627) fusion proteins were also prep'd. and tested for their  
CC efficacy as vaccines against P. aeruginosa infection.  
XX  
SQ Sequence 63 AA;

Query Match 3.1%; Score 7; DB 17; Length 63;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 MLEKASR 119  
Db 56 mlekasr 62

RESULT 14  
R05708  
ID R05708 standard; protein: 83 AA.  
XX  
AC R05708;  
XX  
DT 16-AUG-1990 (first entry)  
XX  
DE Pseudomonas aeruginosa outer membrane lipoprotein I.  
XX  
KW Pseudomonas aeruginosa outer membrane lipoprotein I; vaccine; ss.  
XX  
FH Key Location/Qualifiers

```

FT sig_peptide 1..22
XX
OS Pseudomonas aeruginosa ATCC 33354.
XX
PN EP357024-A.
XX
PD 07-MAR-1990.
XX
PF 30-AUG-1989; 89EP-0115992.
XX
PR 01-SEP-1988; 88DE-3829616.
XX
PA (BEHW ) BEHRINGWERKE AG.
XX
PI Duchene M, von Specht U, Domdey H;
XX WPI; 1990-068794/10.
DR N-PSDB; Q03507.
XX
XX Pseudomonas aeruginosa outer membrane lipoprotein - useful for vaccine
production.
XX
XX Example 5; Page 4; 5pp; German.
XX
XX OMPI and its fragments are useful for the production of vaccines and
XX antiserum against Pseudomonas infections. The DNA and antibodies are
XX useful for diagnostic purposes.
XX
SQ Sequence 83 AA;

Query Match 3.1%; Score 7; DB 11; Length 83;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119
DB 76 mlekasr 82

RESULT 15
B58516
ID B58516 standard; Protein; 130 AA.
XX
AC B58516;
XX
14-MAR-2001 (first entry)
XX
Lung cancer associated polypeptide sequence SEQ ID 854.
XX
Human; lung cancer associated protein; neuroprotective; cytostatic;
XX cardioactive; immunomodulatory; muscular active; vulnerary;
XX gastrointestinal; nephrotropic; antiinfective; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disease; reproductive;
XX proliferative disorder; wound healing; infectious disease.
XX
OS Homo sapiens.
XX
WO200055180-A2.
XX
21-SEP-2000.
XX
08-MAR-2000; 2000WO-US05918.
XX
12-MAR-1999; 99US-0124270.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
Ruben SM;
XX
WPI; 2000-587514/55.
DR N-PSDB; F18392.

```

```

XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX Claim 11; Page 1391-1392; 1425pp; English.
XX
CC Polynucleotide sequences F17982 - F18424 encode human lung cancer
CC associated proteins represented in B58106 - B58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive; and
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences F18425 - F18433 and peptide B58549 are used in
CC the course of the invention for the identification and characterisation
XX of the polynucleotide and protein sequences.
XX
SQ Sequence 130 AA;

Query Match 3.1%; Score 7; DB 21; Length 130;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 LSAMRFL 103
DB 1 lsamrfl 7

Search completed: April 19, 2001, 12:58:35
Job time: 113 sec

```



us-09-441-723-1.rag

Fri Apr 20 10:52:35 2001



Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPLRTVELFVILSPYSLGFEILCRQYQNIWNIQLRPSLITGIMKDSGNKPPGLLPR 61  
 DB 1 GPLRTVELFVILSPYSLGFEILCRQYQNIWNIQLRPSLITGIMKDSGNKPPGLLPR 60

QY 62 KGLYMANDLKLRLHLLQIPIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
 DB 61 KGLYMANDLKLRLHLLQIPIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 120

QY 122 WMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIPATPKVKNQKLEKTEAAACRYGAF 181  
 DB 121 WMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIPATPKVKNQKLEKTEAAACRYGAF 180

QY 182 GLPTTAVHVDGQTHMLFSGDRMELLAHLGKWKMGPIPPAVNARL 226  
 DB 181 GLPTTAVHVDGQTHMLFSGDRMELLAHLGKWKMGPIPPAVNARL 225

RESULT 2  
 GTK1\_RAT STANDARD; PRT: 225 AA.

ID GTK1\_RAT  
 AC P24473; 009034;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GLUTATHIONE S-TRANSFERASE, MITOCHONDRIAL (EC 2.5.1.18) (GST 13-13)  
 DE GLUTATHIONE S-TRANSFERASE SUBUNIT 13 (GST CLASS-KAPPA).  
 GN GSTK1 OR GSTK1-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=97079244; PubMed=8920976;  
 RA Pemble S.E., Wardle A.F., Taylor J.B.;  
 RT "Glutathione S-transferase class kappa: characterization by the  
 RT cloning of rat mitochondrial GST and identification of a human  
 RT homologue.";  
 RL Biochem. J. 319:749-754(1996).  
 RN [2]  
 RP SEQUENCE OF 1-33.  
 RX MEDLINE=91354194; PubMed=1883325;  
 RA Harris M.J., Meyer D.J., Coles B., Ketterer B.;  
 RT "A novel glutathione transferase (13-13) isolated from the matrix of  
 RT rat liver mitochondria having structural similarity to class theta  
 RT enzymes.";  
 RL Biochem. J. 278:137-141(1991).  
 CC -1- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC  
 CC ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT.  
 CC -1- CATALYTIC ACTIVITY: RX + GLUTATHIONE - HX + R-S-GLUTATHIONE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; S83436; AAB50831.1; -;  
 PIR; S17164; S17164.  
 KW Transferase; Mitochondrion.  
 FT INIT MET 0 0  
 FT CONFLICT 1 1 G -> C (IN REF. 2).  
 SQ SEQUENCE 225 AA; 25362 MW; FC895B730655E0C9 CRC64;

Query Match 7.1%; Score 16; DB 1; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-09;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ELFYDVLSPYSWLGFE 24  
 DB 8 ELFYDVLSPYSWLGFE 23

RESULT 3  
 RECA\_AQUY STANDARD; PRT: 348 AA.

ID RECA\_AQUY  
 AC P33542;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE RECA PROTEIN (RECOMBINASE A).  
 GN RECA.  
 OS Aquifex pyrophilus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=2714;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KOL5A;  
 RX MEDLINE=95014407; PubMed=7929298;  
 RA Wetmur J.G., Wong D.M., Ortiz B., Tong J., Reichert F., Gelfand D.H.;  
 RT "Cloning, sequencing, and expression of RecA proteins from three  
 RT distantly related thermophilic eubacteria.";  
 RL J. Biol. Chem. 269:25928-25935(1994).  
 CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF  
 CC SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED  
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF  
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEKA CAUSING  
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L23135; AAA67702.1; -;  
 DR HSPF; P03017; 2REB.  
 DR InterPro; IPR001553; -;  
 DR Pfam; PF00154; RECA; 1.  
 DR PRINTS; PR00142; RECA.  
 DR PROSITE; PS00321; RECA\_1; 1.  
 DR PROSITE; PS0162; RECA\_2; 1.  
 DR PROSITE; PS0163; RECA\_3; 1.  
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.  
 FT NP\_BIND 71 78 ATP (BY SIMILARITY).  
 SQ SEQUENCE 348 AA; 37912 MW; 31E93930989739ED CRC64;

Query Match 3.5%; Score 8; DB 1; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 LEHPEMLE 115  
 DB 315 LEHPEMLE 322

RESULT 4  
 Y4OA\_RHISN STANDARD; PRT: 593 AA.

ID Y4OA\_RHISN  
 AC P55586;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:58:57 ; Search time 10.12 seconds  
(without alignments)  
764.994 Million cell updates/sec

Title: US-09-441-723-1

Perfect score: 226

Sequence: 1 MGPLPRTVELFYDVLSPYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225	99.6	225	1	GTK1_HUMAN
2	16	7.1	225	1	GTK1_RAT
3	16	3.5	348	1	RECA_AQUYP
4	8	3.5	593	1	Y40A_RHISN
5	8	3.5	814	1	SEFC_SALEN
6	7	3.1	71	1	RL29_RICPR
7	7	3.1	83	1	MULI_PSEAE
8	7	3.1	131	1	PRO1_PHLPR
9	7	3.1	131	1	PRO2_PHLPR
10	7	3.1	131	1	PRO3_PHLPR
11	7	3.1	133	1	PROF_BETVE
12	7	3.1	133	1	PROF_MERAN
13	7	3.1	202	1	RS2_PYRHO
14	7	3.1	204	1	TD53_HUMAN
15	7	3.1	204	1	TD53_MOUSE
16	7	3.1	354	1	YHL4_YEAST
17	7	3.1	547	1	DDR2_PSEAE
18	7	3.1	854	1	DDR2_MOUSE
19	7	3.1	855	1	DDR2_HUMAN
20	7	3.1	883	1	YHL6_YEAST
21	7	3.1	902	1	GLR4_HUMAN
22	7	3.1	902	1	GLR4_RAT
23	7	3.1	1875	1	MLP1_YEAST
24	6	2.7	67	1	Y158_RICPR
25	6	2.7	71	1	HPIS_PARSF
26	6	2.7	88	1	RS16_BACST
27	6	2.7	93	1	RL23_HELPY
28	6	2.7	118	1	FHPC_PORPU
29	6	2.7	127	1	NB4M_BOVIN
30	6	2.7	127	1	NB4M_HUMAN
31	6	2.7	130	1	RS11_XANCP
32	6	2.7	131	1	MV47_HUMAN
33	6	2.7	132	1	RL14_HALHA

34	6	2.7	132	1	RL14_HALMA
35	6	2.7	140	1	YB60_HELPY
36	6	2.7	141	1	RL13_HELPY
37	6	2.7	141	1	YB60_HELPJ
38	6	2.7	153	1	XJGD_HAEIN
39	6	2.7	160	1	YFJC_ECOLI
40	6	2.7	165	1	Y400_SYNP2
41	6	2.7	171	1	OGT_ECOLI
42	6	2.7	171	1	OGT_SALTY
43	6	2.7	173	1	RL9_BORBU
44	6	2.7	184	1	YPS2_SYNP2
45	6	2.7	186	1	YG61_METJA

## ALIGNMENTS

RESULT	1
GTK1_HUMAN	
ID	GTK1_HUMAN STANDARD; PRT; 225 AA.
AC	O9Y2Q3; Q9P1S4;
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	GLUTATHIONE S-TRANSFERASE, MITOCHONDRIAL (EC 2.5.1.18) (GST 13-13)
DE	(GLUTATHIONE S-TRANSFERASE SUBUNIT 13) (GST CLASS-KAPPA) (HDCMD47P).
GN	GSTK1
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Blood;
RX	MEDLINE=20499367; PubMed=11042152;
RA	Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA	Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA	Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.,
RT	"Cloning and functional analysis of cDNAs with open reading frames for
RT	300 previously undefined genes expressed in CD34+ hematopoietic
RT	stem/progenitor cells."
RT	Genome Res. 10:1546-1560(2000).
RP	[2]
RP	SEQUENCE FROM N.A.
RA	Zhao Z., Huang X., Li N., Zhu X., Cao X.;
RA	"A novel gene from human dendritic cell."
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC	- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC
CC	ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT (BY SIMILARITY).
CC	- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
CC	- SUBUNIT: HOMODIMER (BY SIMILARITY)
CC	- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC	- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; AF070657; AAD20963.1; -
DR	EMBL; AF068287; AAF65506.1; -
KW	Transferase; Mitochondrion.
FT	INIT_MET 0 0 BY SIMILARITY.
FT	CONFLICT 178 178 G -> R (IN REF. 2).
FT	CONFLICT 219 219 P -> S (IN REF. 2).
SQ	SEQUENCE 225 AA; 25365 MW; FE91A5EE0F0B0BA1 CRC64;

Query Match 99.6%; Score 225; DB 1; Length 225;  
Best Local Similarity 100.0%; Pred. No. 1.1e-221;

Db 2 NDCLKLR 8

RESULT 7

MULTI\_PSEAE STANDARD; PRT; 83 AA.

AC P11221.

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE MAJOR OUTER MEMBRANE LIPOPROTEIN PRECURSOR (MUREIN-LIPOPROTEIN)

DE (LIPOPROTEIN I).

GN OPRI OR PA2853.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OC NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ISOLATE PA2;

RX MEDLINE=89313394; PubMed=2473376;

RA Cornells P., Boula A., Belardi A., Guyonvarch A., Kammerer B.,

RA Hannaert V., Hubert J.-C.;

RT "Cloning and analysis of the gene for the major outer membrane

RT lipoprotein from Pseudomonas aeruginosa.;"

RL Mol. Microbiol. 3:421-428(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=89327122; PubMed=2502533;

RA Duchene M., Barron C., Schweizer A., von Sprecht B.-U., Domdey H.;

RT "Pseudomonas aeruginosa outer membrane lipoprotein I gene: molecular

RT cloning, sequence, and expression in Escherichia coli.;"

RL J. Bacteriol. 171:4130-4137(1989).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN=ATCC 15692 / PA01;

RX MEDLINE=92268853; PubMed=1588307;

RA Saint-Onge A., Roneyer F., Lebel P., Masson L., Brousseau R.;

RT "Specificity of the Pseudomonas aeruginosa PA01 lipoprotein I gene as

RT a DNA probe and PCR target region within the Pseudomonadaceae.;"

RN J. Gen. Microbiol. 138:733-741(1992).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.;"

RL Nature 406:959-964(2000).

CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID

CC ANCHOR.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X13748; CAA32013.1; -.

DR EMBL; M25761; AAA35880.1; -.

DR EMBL; X58714; CAA41550.1; -.

DR EMBL; A07695; CAA00707.1; -.

DR EMBL; AE004712; AAG06241.1; -.

DR FIR; S04834; S04834.

DR PIR; A33854; A33854.

DR PIR; A44834; A44834.

DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Signal.

FT SIGNAL 1 19

FT CHAIN 20 83 MAJOR OUTER MEMBRANE LIPOPROTEIN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE.

SQ SEQUENCE 83 AA; 8835 MW; E87F52B86B04DBA4 CRC64;

Query Match 3.1%; Score 7; DB 1; Length 83;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 MLEKASR 119

Db 76 MLEKASR 82

RESULT 8

PROL\_PHLPR

ID PROL\_PHLPR STANDARD; PRT; 131 AA.

AC P35079;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE PROFILIN 1 (POLLEN ALLERGEN PHL P 11) (PHL P XI).

GN PRO1 OR PHLPXI.

OS Phleum pratense (Common timothy).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; Phleum.

OX NCBI\_TaxID=15957;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pollin;

RX MEDLINE=94168560; PubMed=8123000;

RA Valenta R., Ball T., Vrtala S., Duchene M., Kraft D., Scheiner O.;

RT "cDNA cloning and expression of timothy grass (Phleum pratense)

RT pollen profilin in Escherichia coli: comparison with birch pollen

RT profilin.;"

RL Biochem. Biophys. Res. Commun. 199:106-118(1994).

CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE

CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE

CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW

CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF

CC IP3 AND DG.

CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC

CC ACTIN IN A 1:1 RATIO.

CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.

CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X77583; CAA54686.1; -.

DR PIR; JC2080; JC2080.

DR PIR; S42023; S42023.

DR HSSP; P25816; 1CQA.

DR InterPro; IPR002097; -.

DR Pfam; PF00235; profilin; 1.

DR PRINTS; PR00392; PROFILIN.

DR PROSITE; PS00414; PROFILIN; 1.

KW Actin-binding; Cytoskeleton; Allergen; Multigene family.

SQ SEQUENCE 131 AA; 14235 MW; 1B97A7FCD339714C CRC64;

Query Match 3.1%; Score 7; DB 1; Length 131;

Best Local Similarity 100.0%; Pred. No. 8.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 65.2 KDA PROTEIN Y40A.  
GN Y40A.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
RL Nature 387:394-401(1997).  
CC -1- SIMILARITY: SOME, TO E.COLI PLASMID PMCCC7 MICROCIN C7  
CC BIOSYNTHESIS PROTEIN MCCB.  
CC -1- SIMILARITY: TO Y40C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000088; AAB91794.1; -  
DR InterPro; IPR000594; -  
DR Pfam; PF00899; ThiF\_family; 1.  
KW Hypothetical protein; Plasmid.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 593 AA; 65216 MW; B6E248236AC9F68C CRC64;  
  
Query Match 3.5%; Score 8; DB 1; Length 593;  
Best Local Similarity 100.0%; Pred. No. 3.4; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0;  
  
QY 138 ILAAAEKA 145  
Db 242 ILAAAEKA 249  
|||||||  
  
RESULT 5  
SEFC\_SALEN STANDARD; PRT; 814 AA.  
AC P33388;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE OUTER MEMBRANE USHER PROTEIN SEFC PRECURSOR.  
GN SEFC.  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=27655-3B;  
RX MEDLINE=93239677; PubMed=8097515;  
RA Clouthier S.C., Mueller K.-H., Doran J.L., Collinson S.K., Kay W.W.;  
RT "Characterization of three fimbrial genes, sefABC, of Salmonella  
RT enteritidis.";  
RL J. Bacteriol. 175:2523-2533(1993).  
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE SEFA  
CC FIMBRIAL SUBUNIT.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L11010; AAA27221.1; -  
DR PIR; C40618; C40618.  
DR InterPro; IPR000015; -  
DR Pfam; PF00577; Usher; 1.  
DR PROSITE; PS01151; FIMBRIAL\_USHER; FALSE NEG.  
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 814 OUTER MEMBRANE USHER PROTEIN SEFC.  
FT DISULFID 792 813 POTENTIAL.  
SQ SEQUENCE 814 AA; 90324 MW; AE7C9D35C3FA0EB CRC64;  
  
Query Match 3.5%; Score 8; DB 1; Length 814;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 166 NOLKETTE 173  
Db 60 NOLKETTE 67  
|||||||  
  
RESULT 6  
RL29\_RICPR STANDARD; PRT; 71 AA.  
ID RL29\_RICPR  
AC Q9ZCR3;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 50S RIBOSOMAL PROTEIN L29.  
GN RPMC OR RP651.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MADRID E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria.";  
RL Nature 396:133-140(1998).  
CC -1- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ235272; CAA15091.1; -  
DR InterPro; IPR001854; -  
DR Pfam; PF00831; Ribosomal\_L29; 1.  
DR PROSITE; PS00579; RIBOSOMAL\_L29; 1.  
KW Ribosomal protein.  
SQ SEQUENCE 71 AA; 8406 MW; D3D06C34DA617C3E CRC64;  
  
Query Match 3.1%; Score 7; DB 1; Length 71;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 68 NDKLRL 74  
|||||||

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AP000006; BAA30741.1; ALT\_INIT.  
DR InterPro: IPR001865;  
DR Pfam: PF00318; Ribosomal\_S2; 1.  
DR PRINTS: P00395; RIBOSOMAL\_S2.  
DR PROSITE: PS00962; RIBOSOMAL\_S2\_1; 1.  
DR PROSITE: PS00963; RIBOSOMAL\_S2\_2; 1.  
KW Ribosomal protein.  
SQ SEQUENCE 202 AA; 23022 MW; 686F837D7AFEB57 CRC64;

Query Match 3.1%; Score 7; DB 1; Length 202;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 EPQSILA 140  
DB 63 EPQSILA 69

RESULT 14  
TD53\_HUMAN  
ID TD53\_HUMAN STANDARD; PRT; 204 AA.  
AC Q16890;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TUMOR PROTEIN D53 (HD53) (D52-LIKE 1).  
GN TPDS2L1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Breast carcinoma;  
RA Byrne J.A., Mattei M.-G., Basset P.;  
RT "Definition of the tumor protein D52 (TPD52) gene family through  
RT cloning of D52 homologues in human (hd53) and mouse (md52).";  
RL Genomics 35:523-532(1996).  
RN [2]  
RP INTERACTIONS.  
RX PubMed-9484778;  
RA Byrne J.A., Nourse C.R., Basset P., Gunning P.;  
RT "Identification of homo- and heteromeric interactions between members  
RT of the breast carcinoma-associated D52 protein family using the yeast  
RT two-hybrid system.";  
RL Oncogene 16:873-881(1998).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY.  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U44427; AAB40894.1;  
DR EMBL: U44428; AAB40895.1;  
DR MIM: 604069;  
KW Coiled coil.  
FT DOMAIN 22 73 COILED COIL (POTENTIAL).  
SQ SEQUENCE 204 AA; 22449 MW; 6B3C336D5C0653C9 CRC64;

Query Match 3.1%; Score 7; DB 1; Length 204;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 QAQGLLE 157  
DB 4 QAQGLLE 10

RESULT 15  
TD53\_MOUSE  
ID TD53\_MOUSE STANDARD; PRT; 204 AA.  
AC O54818;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TUMOR PROTEIN D53 (MD53) (D52-LIKE 1).  
GN TPDS2L1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=97001154; PubMed=8812487;  
RA Byrne J.A., Mattei M.-G., Basset P.;  
RT "Definition of the tumor protein D52 (TPD52) gene family through  
RT cloning of D52 homologues in human (hd53) and mouse (md52).";  
RL Genomics 35:523-532(1996).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF004428; AAC98476.1;  
DR MGD; MGI:1298386; Tpd52l1.  
KW Coiled coil.  
FT DOMAIN 22 73 COILED COIL (POTENTIAL).  
SQ SEQUENCE 204 AA; 22515 MW; BAA49DAC7B7F6BE8 CRC64;

Query Match 3.1%; Score 7; DB 1; Length 204;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 QAQGLLE 157  
DB 4 QAQGLLE 10

Search completed: April 19, 2001, 13:00:52  
Job time: 115 sec





QY 44 ITGIMKD 50  
|111111|  
Db 47 ITGIMKD 53

RESULT 9  
PRO3\_PHLPR STANDARD: PRT: 131 AA.  
AC 024650;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PROFILIN 2/4 (POLLEN ALLERGEN PHL P 11) (PHL P XI).  
GN PRO2 AND PRO4.  
OS Phleum pratense (Common timothy).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; Phleum.  
OX NCBI\_TaxID=15957;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pollen;  
RA Asturias J.A., Arilla M.C., Bartolome B., Martinez J., Martinez A.,  
RA Palacios R.;  
RT "Sequence polymorphism and structural analysis of timothy grass  
pollen profilin allergen (Phl p 11)".  
RL Biochim. Biophys. Acta 1352:253-257(1997).  
CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE  
POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW  
CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
IP3 AND DG.  
CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
ACTIN IN A 1:1 RATIO.  
CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.  
CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y09456; CAA70608.1; -;  
DR EMBL; Y09456; CAA70608.1; -;  
DR InterPro; IPR002097; -;  
DR Pfam; PF00235; profilin; 1.  
DR PRINTS; PR00392; PROFILIN.  
DR PROSITE; PS00414; PROFILIN; 1.  
KW Actin-binding; Cytoskeleton; Allergen; Multigene family.  
SQ SEQUENCE 131 AA; 14150 MW; 1B97B9DA084AE14C CRC64;

Query Match 3.1%; Score 7; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
|111111|  
Db 47 ITGIMKD 53

RESULT 10  
PRO3\_PHLPR STANDARD: PRT: 131 AA.  
AC 024282;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PROFILIN 3 (POLLEN ALLERGEN PHL P 11) (PHL P XI).  
GN PRO3.

OS Phleum pratense (Common timothy).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; Phleum.  
OX NCBI\_TaxID=15957;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pollen;  
RA Asturias J.A., Arilla M.C., Bartolome B., Martinez J., Martinez A.,  
RA Palacios R.;  
RT "Sequence polymorphism and structural analysis of timothy grass  
pollen profilin allergen (Phl p 11)".  
RL Biochim. Biophys. Acta 1352:253-257(1997).  
CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE  
POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW  
CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
IP3 AND DG.  
CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
ACTIN IN A 1:1 RATIO.  
CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.  
CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y09457; CAA70609.1; -;  
DR EMBL; Y09457; CAA70609.1; -;  
DR InterPro; IPR002097; -;  
DR Pfam; PF00235; profilin; 1.  
DR PRINTS; PR00392; PROFILIN.  
DR PROSITE; PS00414; PROFILIN; 1.  
KW Actin-binding; Cytoskeleton; Allergen; Multigene family.  
SQ SEQUENCE 131 AA; 14164 MW; CD9A319020E0C7C8 CRC64;

Query Match 3.1%; Score 7; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
|111111|  
Db 47 ITGIMKD 53

RESULT 11  
PROF\_BETVE STANDARD: PRT: 133 AA.  
AC P25816;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PROFILIN (POLLEN ALLERGEN BET V 2) (BET V II).  
GN BETVII.  
OS Betula verrucosa (White birch) (Betula pendula).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotids I;  
OC Fagales; Betulaceae; Betula.  
OX NCBI\_TaxID=3505;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pollen;  
RA Valenta R., Duchene M., Pottenburger K., Sillaber C., Valent P.,  
RA Bettelheim P., Breitenbach M., Rumpold H., Kraft D., Scheiner O.;  
RT "Identification of profilin as a novel pollen allergen; IgE  
autoreactivity in sensitized individuals.";  
RL Science 253:557-560(1991).  
RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=97169442; PubMed=9016715;  
 RA Fedorov A.A., Ball T., Mahoney N.M., Valenta R., Almo S.C.;  
 RT "The molecular basis for allergen cross-reactivity: crystal structure  
 and IgE-epitope mapping of birch pollen profilin.";  
 RL Structure 5:33-45(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=97276803; PubMed=9130496;  
 RA Fedorov A.A., Ball T., Valenta R., Almo S.C.;  
 RT "X-ray crystal structures of birch pollen profilin and Phl p 2.";  
 RL Int. Arch. Allergy Immunol. 113:109-113(1997).  
 RN [4]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=97415424; PubMed=9271223;  
 RA Domke T., Federau T., Schlueter K., Giehl K., Valenta R.,  
 RA Schomburg D., Jockusch B.M.;  
 RT "Birch pollen profilin: structural organization and interaction with  
 poly-(L-proline) peptides as revealed by NMR.";  
 RL FEBS Lett. 411:291-295(1997).  
 CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
 CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE  
 CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW  
 CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
 CC IP3 AND DG.  
 CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
 CC ACTIN IN A 1:1 RATIO.  
 CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.  
 CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M65179; AA16522.1; -;  
 DR PIR: JC2082; JC2082;  
 DR PDB: ICQA; 12-FEB-97.  
 DR InterPro: IPR002097; -;  
 DR Pfam: PF00235; profilin; 1.  
 DR PRINTS: PR00392; PROFILIN.  
 DR PROSITE: PS00414; PROFILIN; 1.  
 KW Actin-binding; Cytoskeleton; Allergen; 3D-structure.  
 SQ SEQUENCE 133 AA; 14253 MW; 9443FC43786E114A CRC64;

Query Match 3.1%; Score 7; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
 DB 49 ITGIMKD 55

RESULT 12  
 PROF\_MERAN STANDARD; PRT; 133 AA.  
 ID AC 049894;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROFILIN (ALLERGEN MER A 1).  
 OS Mercurialis annua (Annual mercury).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;  
 OC Malpighiales; Euphorbiaceae; Mercurialis.  
 OX NCBI\_TaxID=3986;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;  
 RX MEDLINE=98184365; PubMed=9525453;  
 RA Valliverdu A., Asturias J.A., Arilla M.C., Gomez-Bayon N., Martinez A.,  
 RA Martinez J., Palacios R.;  
 RT "Characterization of recombinant Mercurialis annua major allergen Mer  
 a 1 (profilin).";  
 RL J. Allergy Clin. Immunol. 101:363-370(1998).  
 CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE  
 CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE  
 CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW  
 CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF  
 CC IP3 AND DG (BY SIMILARITY).  
 CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC  
 CC ACTIN IN A 1:1 RATIO.  
 CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.  
 CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y13271; CAAT73720.1; -;  
 DR InterPro: IPR002097; -;  
 DR Pfam: PF00235; profilin; 1.  
 DR PRINTS: PR00392; PROFILIN.  
 DR PROSITE: PS00414; PROFILIN; 1.  
 KW Actin-binding; Cytoskeleton; Allergen.  
 SQ SEQUENCE 133 AA; 14300 MW; 06D8C6597217654F CRC64;

Query Match 3.1%; Score 7; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
 DB 49 ITGIMKD 55

RESULT 13  
 RS2\_PVRHO STANDARD; PRT; 202 AA.  
 ID AC 039295;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 30S RIBOSOMAL PROTEIN S2P.  
 GN RPS2P OR PH1629.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 CC -!- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:57:26 ; Search time 15.56 Seconds  
(without alignments)  
998.161 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGPLPRTVELFDVLSVSW.....AHLLEGKMGPIPPAVNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 198801 seqs, 68722935 residues

Word size : 0

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR.67:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	11	4.9	33	S17164	glutathione transf
2	8	3.5	77	C82988	hypothetical prote
3	8	3.5	305	T36056	hypothetical prote
4	8	3.5	348	A55020	recombination prot
5	8	3.5	814	C40618	fimbrial outer mem
6	7	3.1	71	A71671	ribosomal protein
7	7	3.1	83	A33854	outer membrane lip
8	7	3.1	95	F81873	probable membrane
9	7	3.1	131	JC2080	profilin - common
10	7	3.1	133	JC2082	profilin - Europea
11	7	3.1	199	E75637	probable chromosom
12	7	3.1	205	E71042	probable ribosomal
13	7	3.1	212	D70489	hypothetical prote
14	7	3.1	220	G83247	2-keto-3-deoxy-6-p
15	7	3.1	251	T08315	hypothetical prote
16	7	3.1	276	H81931	probable bis(5'-nu
17	7	3.1	279	T36086	probable binding p
18	7	3.1	287	T39197	yeast atp12 protei
19	7	3.1	288	D82781	conserved hypothet
20	7	3.1	347	T43133	hypothetical prote
21	7	3.1	357	T01434	NADPH HC toxin red
22	7	3.1	364	S77360	cbid protein - syn
23	7	3.1	376	B75260	conserved hypothet
24	7	3.1	376	E83604	hypothetical prote
25	7	3.1	396	T27946	hypothetical prote
26	7	3.1	401	F83187	probable pyridoxal
27	7	3.1	458	T34574	hypothetical prote
28	7	3.1	498	H82679	two-component syst
29	7	3.1	547	H83018	dihydrolipoamide a

30	7	3.1	596	2	H81095	conserved hypothet
31	7	3.1	601	2	F81846	hypothetical secre
32	7	3.1	702	2	C82792	methionyl-tRNA syn
33	7	3.1	723	2	T14765	hypothetical prote
34	7	3.1	819	2	I48859	tyro 10 receptor k
35	7	3.1	832	2	S46706	probable transcrip
36	7	3.1	855	2	S42621	protein-tyrosine k
37	7	3.1	884	2	A44839	glutamate receptor
38	7	3.1	902	2	D40170	glutamate RND efflu
39	7	3.1	1036	2	B83466	probable RND prote
40	7	3.1	1525	2	T14961	hypothetical prote
41	7	3.1	1563	2	T28923	myosin-like protei
42	7	3.1	1875	2	S38173	probable non-ribos
43	7	3.1	2352	2	C83229	probable cell-sur
44	7	3.1	3229	2	S27852	probable repressor
45	6	2.7	60	2	H72785	

## ALIGNMENTS

### RESULT 1

S17164  
glutathione transferase (EC 2.5.1.18) l3 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
C:Accession: S17164  
R:Harris, J.M.; Meyer, D.J.; Coles, B.; Ketterer, B.  
Biochem. J. 278, 137-141, 1991  
A:Title: A novel glutathione transferase (l3-l3) isolated from the matrix of rat  
A:Reference number: S17164; MUID:91354194  
A:Accession: S17164  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-33 <HAR>  
C:Keywords: transferase

Query Match 4.9%; Score 11; DB 2; Length 33;  
Best Local Similarity 100.0%; Pred. No. 0.00027;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ELFYDVLSPYS 19  
DB 8 ELFYDVLSPYS 18  
|||||

### RESULT 2

C82988  
hypothetical protein PA5271 [Imported] - Pseudomonas aeruginosa (strain PAO1)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C82988  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.; Adman, S.; Xuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: C82988  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-77 <STO>  
A:Cross-references: GB:AE004939; GB:AE004091; NID:g9951575; PIDN:AAG08656.1; GSPD01  
C:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: PA5271

Query Match 3.5%; Score 8; DB 2; Length 77;  
Best Local Similarity 100.0%; Pred. No. 0.79;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AEQAQGLL 156  
|||||||  
Db 65 AEQAQGLL 72

## RESULT 3

T36056  
hypothetical protein SCD78.17c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
C:Accession: T36056  
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z21595  
A:Accession: T36056  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-305 <SAU>  
A:Cross-references: EMBL:AL034355; PIDN:CAA22222.1; GSPDB:GN00070; SCODEB:SCD78.17c  
C:Genetics:  
A:Gene: SCODEB:SCD78.17c  
C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH10

Query Match 3.5%; Score 8; DB 2; Length 305;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LAAAEKAG 146  
|||||||  
Db 59 LAAAEKAG 66

## RESULT 4

A55020  
recombination protein recA - Aquifex pyrophilus

N:Alternate names: recombinase A  
C:Species: Aquifex pyrophilus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 26-Aug-1999  
C:Accession: A55020  
R:Wetmur, J.G.; Wong, D.M.; Ortiz, B.; Tong, J.; Reichert, F.; Gelfand, D.H.  
J. Biol. Chem. 269, 25928-25935, 1994

A:Title: Cloning, sequencing, and expression of RecA proteins from three distantly related  
A:Reference number: A55020; MUID:95014407

A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA

A:Residues: 1-348 <WET>  
A:Cross-references: GB:L23135; NID:g349396; PIDN:AAA67702.1; PID:g349397

C:Superfamily: recombination protein recA  
C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; P-loop; SOS response  
F:71-78/Region: nucleotide-binding motif A (P-loop)  
F:145-150/Region: nucleotide-binding motif B  
F:77/Binding site: ATP (Lys) #status predicted

Query Match 3.5%; Score 8; DB 2; Length 348;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 LEHPEMLE 115  
|||||||  
Db 315 LEHPEMLE 322

## RESULT 5

C40618

fimbrial outer membrane protein homolog SefC - Salmonella enteritidis

C:Species: Salmonella enteritidis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C40618  
R:Clouthier, S.C.; Muller, K.H.; Doran, J.L.; Collinson, S.K.; Kay, W.W.

J. Bacteriol. 175, 2523-2533, 1993  
A:Title: Characterization of three fimbrial genes, sefABC, of Salmonella enteritidis  
A:Reference number: A40618; MUID:93239677

A:Contents: 27655-3b  
A:Accession: C40618  
A:Status: preliminary  
A:Molecule type: DNA

A:Cross-references: GB:L11010; NID:g310649; PIDN:AAA27221.1; PID:g310650  
A:Note: sequence extracted from NCBI backbone (NCBIN:130387, NCBIP:130397)  
C:Superfamily: outer membrane usher protein fimb  
C:Keywords: membrane protein

Query Match 3.5%; Score 8; DB 1; Length 814;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 NOLKETTE 173  
|||||||  
Db 60 NOLKETTE 67

## RESULT 6

A71671

ribosomal protein L29 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: A71671  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsm  
Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondr  
A:Reference number: A71630; MUID:99039499

A:Accession: A71671  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-71 <AND>

A:Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA15091.1; PID:  
A:Experimental source: strain Madrid E

C:Genetics:  
A:Gene: rpmC; RP651

Query Match 3.1%; Score 7; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 NDLKLLR 74  
|||||||  
Db 2 NDLKLLR 8

## RESULT 7

A33854

outer membrane lipoprotein I precursor - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa  
C:Date: 27-Feb-1990 #sequence\_revision 27-Feb-1990 #text\_change 31-Dec-2000  
C:Accession: A33854; S04834; A44834; A83288  
R:Duchene, M.; Barron, C.; Schweizer, A.; von Specht, B.U.; Domdey, H.  
J. Bacteriol. 171, 4130-4137, 1989

A:Title: Pseudomonas aeruginosa outer membrane lipoprotein I gene: molecular cloning  
A:Reference number: A33854; MUID:89327122

A:Accession: A33854

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-83 <DUC>

A:Cross-references: GB:M25761; NID:g151334; PIDN:AAA25880.1; PID:g151335  
R:Corneilis, P.; Boula, A.; Belarbi, A.; Guyonvarch, A.; Kammerer, B.; Hannaert, V.;  
Mol. Microbiol. 3, 421-428, 1989

A:Title: Cloning and analysis of the gene for the major outer membrane lipoprotein  
A:Reference number: S04834; MUID:89313294

A:Accession: S04834

A:Molecule type: DNA

RESULT 10  
JC2082  
profilin - European white birch

A;Residues: 1-199 <WHI>

A:Cross-references: GB:AE001827; NID:g6460959; PIDN:AAF12679.1; PID:g6460976; TIGR:DR000  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRC0025  
 A:Map position: plasmid  
 A:Genome: plasmid  
 A:Note: plasmid CPI

Query Match 3.1%; Score 7; DB 2; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 LQLRPSL 43  
 Db 130 LQLRPSL 136  
 |||||

## RESULT 12

E71042  
 probable ribosomal protein S2 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 18-Aug-2000  
 C:Accession: E71042  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: E71042  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-205 <KAW>  
 A:Cross-references: GB:AF000006; NID:g3236133; PIDN:BAA30741.1; PID:g3258058  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH1629  
 C:Superfamily: yeast ribosomal protein S1.e

Query Match 3.1%; Score 7; DB 2; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 EPQSILA 140  
 Db 66 EPQSILA 72  
 |||||

## RESULT 13

D70489  
 hypothetical protein aq\_2203 - Aquifex aeolicus  
 C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
 C:Accession: D70489  
 R:DeKert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
 V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666  
 A:Accession: D70489  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-212 <AQF>  
 A:Cross-references: GB:AE000777; NID:g2984377; PIDN:AAC07907.1; PID:g2984389; GB:AE00065  
 A:Experimental source: strain VF5  
 C:Genetics:  
 A:Gene: aq\_2203  
 C:Superfamily: Aquifex aeolicus hypothetical protein aq\_2203

Query Match 3.1%; Score 7; DB 2; Length 212;

Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 LEKGSLS 98  
 Db 156 LEKGSLS 162  
 |||||

## RESULT 14

G83247  
 2-keto-3-deoxy-6-phosphogluconate aldolase PA3181 [imported] - Pseudomonas aerugin  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83247  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: G83247  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-220 <STO>  
 A:Cross-references: GB:AE004742; GB:AE004091; NID:g9949294; PIDN:AAG06569.1; GSPDB  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA3181  
 C:Superfamily: 2-dehydro-3-deoxyphosphogluconate aldolase

Query Match 3.1%; Score 7; DB 2; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 AAEEKAG 146  
 Db 84 AAEEKAG 90  
 |||||

## RESULT 15

T08315  
 hypothetical protein H1185 [imported] - Halobacterium sp. (strain NRC-1) plasmid p  
 C:Species: Halobacterium sp.  
 A:Variety: strain NRC-1  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 03-Nov-2000  
 C:Accession: T08315  
 R:Ng, W.V.; Clufo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall  
 Genome Res. 8, 1131-1141, 1998  
 A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid  
 A:Reference number: 216408; MUID:99063795  
 A:Accession: T08315  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-251 <DAS>  
 A:Cross-references: EMBL:AF016485; NID:g2822278; PID:g2822376; HALOSP:H1185  
 A:Experimental source: strain NRC-1  
 C:Genetics:  
 A:Gene: HALOSP:H1185  
 A:Genome: plasmid pNRC100  
 C:Superfamily: Halobacterium plasmid pNRC100 hypothetical protein H1185

Query Match 3.1%; Score 7; DB 2; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

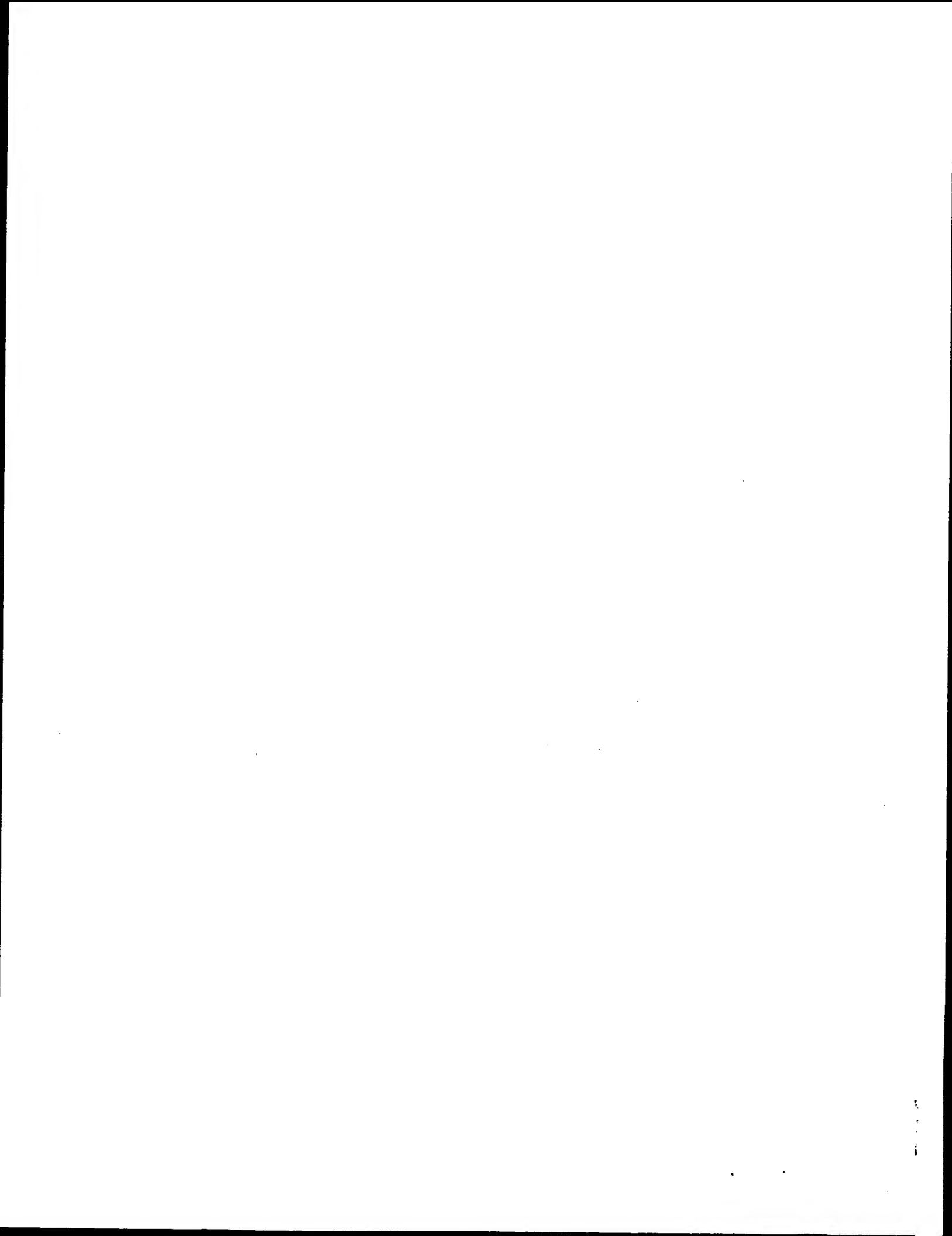
Qy 204 ELLAHL 210  
 Db 64 ELLAHL 70  
 |||||

Fri Apr 20 10:52:39 2001

us-09-441-723-1.rpr

Page 5

Search completed: April 19, 2001, 12:59:14  
Job time: 108 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM.protein - protein search, using sw model

Run on: April 19, 2001, 12:57:21 ; Search time 14.05 Seconds  
(without alignments)  
309.014 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGPLPRTVELFDVLSYSW.....AHLGKMGWPIPPAYNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 185757 seqs, 19210857 residues

Word size : 0

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	226	100.0	226	3	US-08-978-174-1 Sequence 1, Appli
2	16	7.1	226	3	US-08-978-174-3 Sequence 3, Appli
3	8	3.5	814	1	US-08-233-788A-42 Sequence 42, Appl
4	7	3.1	63	2	US-08-572-447C-9 Sequence 9, Appli
5	7	3.1	131	1	US-07-846-992-11 Sequence 11, Appl
6	7	3.1	131	1	US-08-469-555-11 Sequence 11, Appl
7	7	3.1	133	1	US-07-846-992-3 Sequence 3, Appli
8	7	3.1	133	1	US-08-469-555-3 Sequence 3, Appli
9	7	3.1	204	2	US-08-715-204-1 Sequence 1, Appli
10	7	3.1	204	2	US-08-691-814B-10 Sequence 10, Appl
11	7	3.1	204	3	US-09-162-597-1 Sequence 1, Appli
12	7	3.1	214	2	US-08-572-447C-13 Sequence 13, Appl
13	7	3.1	221	3	US-09-031-563-22 Sequence 22, Appl
14	7	3.1	226	2	US-08-572-447C-15 Sequence 15, Appl
15	7	3.1	291	1	US-08-687-379-14 Sequence 14, Appl
16	7	3.1	291	1	US-08-687-379-16 Sequence 16, Appl
17	7	3.1	685	3	US-09-031-563-21 Sequence 21, Appl
18	7	3.1	854	2	US-08-456-647B-20 Sequence 20, Appl
19	7	3.1	854	2	US-08-237-401A-20 Sequence 20, Appl
20	7	3.1	855	1	US-08-336-343A-4 Sequence 4, Appli
21	7	3.1	902	1	US-07-718-575-8 Sequence 8, Appli
22	7	3.1	902	1	US-08-259-164-2 Sequence 2, Appli
23	7	3.1	902	1	US-08-481-206-8 Sequence 8, Appli
24	7	3.1	902	2	US-08-486-289A-8 Sequence 8, Appli
25	7	3.1	1055	3	US-09-031-563-27 Sequence 27, Appl
26	7	3.1	1315	3	US-09-031-563-2 Sequence 2, Appli
27	7	3.1	1315	3	US-09-031-563-25 Sequence 25, Appl

28 6 2.7 16 1 US-08-391-000-20 Sequence 20, Appli  
29 6 2.7 16 2 US-08-741-931-20 Sequence 20, Appli  
30 6 2.7 16 4 US-08-847-844A-47 Sequence 47, Appli  
31 6 2.7 28 2 US-08-451-371-1 Sequence 1, Appli  
32 6 2.7 28 2 US-08-451-371-2 Sequence 2, Appli  
33 6 2.7 28 2 US-08-451-371-3 Sequence 3, Appli  
34 6 2.7 28 2 US-08-451-371-4 Sequence 4, Appli  
35 6 2.7 28 2 US-08-451-371-5 Sequence 5, Appli  
36 6 2.7 101 2 US-08-710-749-7 Sequence 7, Appli  
37 6 2.7 128 2 US-08-946-528-5 Sequence 9, Appli  
38 6 2.7 128 2 US-08-946-528-5 Sequence 9, Appli  
39 6 2.7 141 1 US-08-442-063A-36 Sequence 36, Appli  
40 6 2.7 157 2 US-08-164-292B-4 Sequence 4, Appli  
41 6 2.7 157 3 US-08-845-623-4 Sequence 4, Appli  
42 6 2.7 157 3 US-08-815-927-4 Sequence 4, Appli  
43 6 2.7 169 2 US-08-460-309-20 Sequence 20, Appli  
44 6 2.7 169 2 US-08-125-077-20 Sequence 20, Appli  
45 6 2.7 174 2 US-08-683-262B-40 Sequence 40, Appli

## ALIGNMENTS

RESULT 1  
US-08-978-174-1  
; Sequence 1, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADTUT04  
; CLONE: 1554593  
US-08-978-174-1

Query Match 100.0%; Score 226; DB 3; Length 226;  
Best Local Similarity 100.0%; Pred. No. 4.5e-223;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPLRPTVELFDVLSYSPYSLGFEILCRQYQINWINQLRPSLTGIMKDSGNKPPGLLP 60  
Db 1 MGPLRPTVELFDVLSYSPYSLGFEILCRQYQINWINQLRPSLTGIMKDSGNKPPGLLP 60  
Qy 61 RKGLYMANDLKLRLHILQIPHPKDFLSVNLKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
Db 61 RKGLYMANDLKLRLHILQIPHPKDFLSVNLKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
Qy 121 LMRVMSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKQKETEAAACRYGA 180  
Db 121 LMRVMSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKQKETEAAACRYGA 180  
Qy 181 FGLPTVAHVDCQTHMLFSGDRMELLAHLGKWMGPIPPAVNARL 226  
Db 181 FGLPTVAHVDCQTHMLFSGDRMELLAHLGKWMGPIPPAVNARL 226

RESULT 2

US-08-978-174-3  
; Sequence 3, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Nell C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: ?  
US-08-978-174-3

Query Match 7.1%; Score 16; DB 3; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 9 ELFYDVLSPYSWLGFE 24  
Db 9 ELFYDVLSPYSWLGFE 24

RESULT 3

US-08-233-788A-42  
; Sequence 42, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 814 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-42

Query Match 3.5%; Score 8; DB 1; Length 814;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 NOLKETTE 173  
Db 60 NOLKETTE 67

RESULT 4

US-08-572-447C-9  
; Sequence 9, Application US/08572447C  
; Patent No. 5955090  
; GENERAL INFORMATION:  
; APPLICANT: Knapp, Bernhard  
; APPLICANT: Hungerer, Klaus-Dieter  
; APPLICANT: Broker, Michael  
; APPLICANT: Von Specht, Bernd-Ulrich  
; APPLICANT: Domdey, Horst  
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI  
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.

;;  
;; NUMBER OF SEQUENCES: 15  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &  
;; ADDRESSEE: Dunner  
;; STREET: 1300 I Street, N.W., Suite 700  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-3315  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/572.447C  
;; FILING DATE: 14-DEC-1995  
;; CLASSIFICATION: 424  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: EP 94120023.0  
;; FILING DATE: 16-DEC-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: M. Paul Barker  
;; REGISTRATION NUMBER: 32,013  
;; REFERENCE/DOCKET NUMBER: 05552.1395-00000  
;;  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 408-4000  
;; TELEFAX: (202) 408-4400  
;;  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 63 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
;; US-08-572-447C-9

Query Match 3.1%; Score 7; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 MLEKASR 119  
Db 56 MLEKASR 62

RESULT 5  
US-07-846-992-11  
; Sequence 11, Application US/07846992  
; Patent No. 5583046  
; GENERAL INFORMATION:  
; APPLICANT: Valenta, Rudolf  
; APPLICANT: Duchene, Michael  
; APPLICANT: Pettenburger, Karin  
; APPLICANT: Breitenbach, Michael  
; APPLICANT: Kraft, Dietrich  
; APPLICANT: Rumpold, Helmut  
; APPLICANT: Scheiner, Otto  
; TITLE OF INVENTION: Birch Pollen Allergen Pl4 for Diagnosis  
; TITLE OF INVENTION: and Therapy of Allergic Diseases  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/846.992  
;; FILING DATE: 19920606  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/353.844  
;; FILING DATE: 18-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jones III, Harry C  
;; REGISTRATION NUMBER: 20,280  
;; REFERENCE/DOCKET NUMBER: 6530-011  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 131 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Phleum pratense  
;; FEATURE:  
;; OTHER INFORMATION: Amino acid identity with p14  
; allergen from Betula verrucosa is 77%  
;;  
;; US-07-846-992-11

Query Match 3.1%; Score 7; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 ITGIMKD 50  
Db 47 ITGIMKD 53

RESULT 6  
US-08-469-555-11  
; Sequence 11, Application US/08469555  
; Patent No. 5648242  
; GENERAL INFORMATION:  
; APPLICANT: Valenta, Rudolf  
; APPLICANT: Duchene, Michael  
; APPLICANT: Pettenburger, Karin  
; APPLICANT: Breitenbach, Michael  
; APPLICANT: Kraft, Dietrich  
; APPLICANT: Rumpold, Helmut  
; APPLICANT: Scheiner, Otto  
; TITLE OF INVENTION: Birch Pollen Allergen Pl4 for Diagnosis  
; TITLE OF INVENTION: and Therapy of Allergic Diseases  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469.555  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/846.992  
; FILING DATE: 06-JUN-1992

APPLICATION NUMBER: US/07/353,844  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones III, Harry C  
REGISTRATION NUMBER: 20,280  
REFERENCE/DOCKET NUMBER: 6530-011  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Phleum pratense  
FEATURE:  
OTHER INFORMATION: Amino acid identity with P14  
OTHER INFORMATION: allergen from Betula verrucosa is 77%  
US-08-469-555-11

Query Match 3.1%; Score 7; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
DB 47 ITGIMKD 53

## RESULT 7

US-07-846-992-3  
Sequence 3, Application US/07846992  
Patent No. 5583046  
GENERAL INFORMATION:  
APPLICANT: Valenta, Rudolf  
APPLICANT: Duchene, Michael  
APPLICANT: Pottenburger, Karin  
APPLICANT: Breitenbach, Michael  
APPLICANT: Kraft, Dietrich  
APPLICANT: Rumpold, Helmut  
APPLICANT: Scheiner, Otto  
TITLE OF INVENTION: Birch Pollen Allergen P14 for Diagnosis  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/846,992  
FILING DATE: 19920606  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/353,844  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones III, Harry C  
REGISTRATION NUMBER: 20,280  
REFERENCE/DOCKET NUMBER: 6530-011  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 133 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Betula verrucosa  
FEATURE:  
OTHER INFORMATION: Amino acid sequence identity  
OTHER INFORMATION: with profilin of other organisms is as follows:  
OTHER INFORMATION: 30% with human profilin, 28% with calf and mouse,  
OTHER INFORMATION: 26% with yeast and 25% with Acanthamoeba  
US-07-846-992-3

Query Match 3.1%; Score 7; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50  
DB 49 ITGIMKD 55

## RESULT 8

US-08-469-555-3  
Sequence 3, Application US/08469555  
Patent No. 5648242  
GENERAL INFORMATION:  
APPLICANT: Valenta, Rudolf  
APPLICANT: Duchene, Michael  
APPLICANT: Pottenburger, Karin  
APPLICANT: Breitenbach, Michael  
APPLICANT: Kraft, Dietrich  
APPLICANT: Rumpold, Helmut  
APPLICANT: Scheiner, Otto  
TITLE OF INVENTION: Birch Pollen Allergen P14 for Diagnosis  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,555  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/846,992  
FILING DATE: 06-JUN-1992  
APPLICATION NUMBER: US/07/353,844  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones III, Harry C  
REGISTRATION NUMBER: 20,280  
REFERENCE/DOCKET NUMBER: 6530-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE

## ; INFORMATION FOR SEQ ID NO: 3:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 133 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Betula verrucosa

## ; FEATURE:

; OTHER INFORMATION: Amino acid sequence identity

; OTHER INFORMATION: with profilin of other organisms is as follows:

; OTHER INFORMATION: 30% with human profilin, 28% with calf and mouse,

; OTHER INFORMATION: 26% with yeast and 25% with Acanthamoeba

; US-08-469-555-3

Query Match 3.1%; Score 7; DB 1; Length 133;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ITGIMKD 50

| | | | |

DB 49 ITGIMKD 55

## RESULT 9

US-08-715-204-1

; Sequence 1, Application US/08715204

; Patent No. 5874286

## ; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Au-Young, Janice

; APPLICANT: Goli, Surya K.

; APPLICANT: Hillman, Jennifer.

; APPLICANT: Zweiger, Gary B.

; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: U.S.

; ZIP: 94304

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/715,204

; FILING DATE: Filed Herewith

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0126 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 204 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; IMMEDIATE SOURCE:

; LIBRARY:

; CLONE: Consensus

US-08-715-204-1

Query Match 3.1%; Score 7; DB 2; Length 204;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 QAQGLLE 157

| | | | |

DB 4 QAQGLLE 10

## RESULT 10

US-08-691-814B-10

; Sequence 10, Application US/08691814B

; Patent No. 5981218

## ; GENERAL INFORMATION:

; APPLICANT: Rio, Marie-Christine

; APPLICANT: Tomasetto, Catherine

; APPLICANT: Bassett, Paul

; APPLICANT: Byrne, Jennifer

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful

; TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis

; NUMBER OF SEQUENCES: 124

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox P.L.L.C.

; STREET: 1100 New York Ave, NW, Suite 600

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3934

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/691,814B

; FILING DATE: 31-JUL-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/002,183

; FILING DATE: 09-AUG-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Steffe, Eric K.

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 1383.0090001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2543

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 204 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-691-814B-10

Query Match 3.1%; Score 7; DB 2; Length 204;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 QAQGLLE 157

| | | | |

DB 4 QAQGLLE 10

## RESULT 11

US-09-162-597-1

; Sequence 1, Application US/09162597

; Patent No. 6043343

## ; GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer.  
APPLICANT: Zweiger, Gary B.  
TITLE OF INVENTION: A NOVEL TUMOR PROTEIN  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/162,597  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/715,204  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0126 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 204 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE: Consensus  
US-09-162-597-1

Query Match 3.1% Score 7; DB 3; Length 204;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 QAQGLLE 157  
|||||  
Db 4 QAQGLLE 10

RESULT 12  
US-08-572-447C-13  
; Sequence 13, Application US/08572447C  
; Patent No. 5955090  
; GENERAL INFORMATION:  
; APPLICANT: Knapp, Bernhard  
; APPLICANT: Hungeler, Klaus-Dieter  
; APPLICANT: Broker, Michael  
; APPLICANT: Von Specht, Bernd-Ulrich  
; APPLICANT: Dornay, Horst  
TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI  
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.

COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/572,447C  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 94120023.0  
FILING DATE: 16-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: M. Paul Barker  
REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 05552.1395-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-572-447C-13

Query Match 3.1% Score 7; DB 2; Length 214;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
|||||  
Db 207 MLEKASR 213

RESULT 13  
US-09-031-563-22  
; Sequence 22, Application US/09031563A  
; Patent No. 6022708  
; GENERAL INFORMATION:  
; APPLICANT: Frederic de Sauvage  
; APPLICANT: Arnon Roshenthal  
; TITLE OF INVENTION: Fused  
; FILE REFERENCE: PI272  
; CURRENT APPLICATION NUMBER: US/09/031,563A  
; CURRENT FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 27  
; SEQ ID NO 22  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-031-563-22

Query Match 3.1% Score 7; DB 3; Length 221;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 QLRPSLI 44  
|||||  
Db 66 QLRPSLI 72

RESULT 14  
US-08-572-447C-15  
; Sequence 15, Application US/08572447C  
; Patent No. 5955090  
; GENERAL INFORMATION:  
; APPLICANT: Knapp, Bernhard

```

; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572.447C
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Paul Barker
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-572-447C-15

```

```

Query Match 3.1%; Score 7; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 113 MLEKASR 119
Db 56 MLEKASR 62

```

```

RESULT 15
US-08-687-379-14
; Sequence 14, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ultsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: Wordperfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-687-379-14

```

```

Query Match 3.1%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 37 LQLRPSL 43
Db 132 LQLRPSL 138

```

```

Search completed: April 19, 2001, 12:58:52
Job time: 91 sec

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:58:37 ; Search time 24.77 Seconds  
(without alignments)  
1069.398 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGPLPRTVELFYDVLSPYSW.....AHLGLKWMGPIPPAVNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Capext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: SPTRMBL15.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	226	100.0	226	4 Q9Y2Q3	Q9Y2Q3 homo sapien
2	178	78.8	226	4 Q9P1S4	Q9P1S4 homo sapien
3	8	3.5	283	4 Q9UL42	Q9UL42 homo sapien
4	8	3.5	305	2 Q9ZBY2	Q9ZBY2 streptomyce
5	8	3.5	364	4 Q94959	Q94959 homo sapien
6	8	3.5	392	13 Q9IBF0	Q9IBF0 anguilla ja
7	8	3.5	419	13 Q9PUT7	Q9PUT7 brachydanio
8	8	3.5	426	13 Q9W611	Q9W611 cynops pyrr
9	7	3.1	83	2 Q85409	Q85409 pseudomonas
10	7	3.1	83	2 Q85410	Q85410 pseudomonas
11	7	3.1	83	2 Q85411	Q85411 pseudomonas
12	7	3.1	83	2 Q85412	Q85412 pseudomonas
13	7	3.1	83	2 Q85413	Q85413 pseudomonas
14	7	3.1	83	2 Q85414	Q85414 pseudomonas
15	7	3.1	83	2 Q85415	Q85415 pseudomonas
16	7	3.1	83	2 Q85416	Q85416 pseudomonas
17	7	3.1	83	2 Q85417	Q85417 pseudomonas
18	7	3.1	83	2 Q85418	Q85418 pseudomonas
19	7	3.1	83	2 Q85419	Q85419 pseudomonas

20	7	3.1	83	2	Q85420	pseudomonas
21	7	3.1	83	2	Q85421	pseudomonas
22	7	3.1	83	2	Q85422	pseudomonas
23	7	3.1	83	2	Q85423	pseudomonas
24	7	3.1	83	2	Q85424	pseudomonas
25	7	3.1	83	2	Q85425	pseudomonas
26	7	3.1	83	2	Q85426	pseudomonas
27	7	3.1	83	2	Q85427	pseudomonas
28	7	3.1	83	2	Q85428	pseudomonas
29	7	3.1	83	2	Q85430	pseudomonas
30	7	3.1	83	2	Q85431	pseudomonas
31	7	3.1	83	2	Q85432	pseudomonas
32	7	3.1	83	2	Q85433	pseudomonas
33	7	3.1	83	2	Q85434	pseudomonas
34	7	3.1	83	2	Q85435	pseudomonas
35	7	3.1	83	2	Q85437	pseudomonas
36	7	3.1	83	2	Q85438	pseudomonas
37	7	3.1	83	2	Q85439	pseudomonas
38	7	3.1	83	2	Q85440	pseudomonas
39	7	3.1	83	2	Q85441	pseudomonas
40	7	3.1	83	2	Q85442	pseudomonas
41	7	3.1	83	2	Q85444	pseudomonas
42	7	3.1	83	2	Q85445	pseudomonas
43	7	3.1	122	2	Q9JUY6	Q9JUY6 neisseria m
44	7	3.1	189	2	Q9RHW6	Q9RHW6 methylobaci
45	7	3.1	192	2	Q9RHW5	Q9RHW5 borrelia bu

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	226 AA.
Q9Y2Q3				
ID	Q9Y2Q3			
AC	Q9Y2Q3;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)			
DE	GLUTATHIONE S-TRANSFERASE SUBUNIT 13 HOMOLOG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Mao M., Ye M., Zhang Q., Zhou J., Wu J., Shen Y., Kan L., He K.,			
RA	Gu B., Fu G., Chen S., Chen Z.,			
RT	"Human rGSTK1-1 homolog gene."			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF070657; AAD20963.1; "			
KW	Transferase			
SQ	SEQUENCE 226 AA; 25497 MW; D3FDAFD1533B58A4 CRC64;			

Query Match	100.0%;	Score 226;	DB 4;	Length 226;
Best Local Similarity	100.0%;	Pred. No. 2.9e-221;		
Matches 226;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MGPLPRTVELFYDVLSPYSWLGFELCYRQNIWNINQLRPSLITGIMKDSGNKPPGLLP	60	
Db	1	MGPLPRTVELFYDVLSPYSWLGFELCYRQNIWNINQLRPSLITGIMKDSGNKPPGLLP	60	
QY	61	RKGLYMANDLLRLHLLQIPHPKDFLSVLMKSGLSAMRFLTAVNLEHPEMLEKASRE	120	
Db	61	RKGLYMANDLLRLHLLQIPHPKDFLSVLMKSGLSAMRFLTAVNLEHPEMLEKASRE	120	
QY	121	LWMEVWSRNEDETEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKLNKLTETEAACRYGA	180	
Db	121	LWMEVWSRNEDETEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKLNKLTETEAACRYGA	180	
QY	181	FGLPITVAHVQDQTHLMFGSDRMELLAHLIGEKGWGPPIPPAVNARL	226	
Db	181	FGLPITVAHVQDQTHLMFGSDRMELLAHLIGEKGWGPPIPPAVNARL	226	

```
RESULT 2
Q9P1S4 PRELIMINARY; PRT; 226 AA.
AC Q9P1S4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HDCMD47P.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhao Z., Huang X., Li N., Zhu X., Cao X.;
RT "A novel gene from human dendritic cell.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068287; AAF65506.1; -.
SQ SEQUENCE 226 AA; 25586 MW; D3FDB561533B5A65 CRC64;

Query Match 78.8%; Score 178; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.6e-172;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLTGIMKDSGNKPPGLLP 60
DB 1 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLTGIMKDSGNKPPGLLP 60

QY 61 RKGLYMANDLKLRRHLOIPIHFPRKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120
DB 61 RKGLYMANDLKLRRHLOIPIHFPRKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120

QY 121 LMRVRSRNEDEITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEACRY 178
DB 121 LMRVRSRNEDEITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEACRY 178

RESULT 3
Q9UL42 PRELIMINARY; PRT; 283 AA.
AC Q9UL42;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PARANEOPLASTIC CANCER-TESTIS-BRAIN ANTIGEN (FRAGMENT).
GN MA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gultekin S.H., Voltz R., Rosenfeld M.R., Gerstner E., Eichen J.,
RA Posner J.B., Dalmat J.;
RT "Identification of a novel cancer testis brain antigen using serum
RT antibodies from patients with testicular tumors and paraneoplastic
RT limbic encephalomyelitis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083115; AAF05626.1; -.
FT NON_TER 1
SQ SEQUENCE 283 AA; 32333 MW; E27BA6BCDCD240A4 CRC64;

Query Match 3.5%; Score 8; DB 4; Length 283;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ELLAHLGL 211
DB 59 ELLAHLGL 66
```

```
RESULT 4
Q92BY2 PRELIMINARY; PRT; 305 AA.
AC Q92BY2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 32.5 KDA PROTEIN.
GN SCD78.17C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Elchner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL034355; CA222222.1; -.
DR INTERPRO; IPR000150; -.
DR PFAM; PF00592; DUF3; 1.
KW Hypothetical protein.
SQ SEQUENCE 305 AA; 32463 MW; B8DCCB51C25E3B35 CRC64;

Query Match 3.5%; Score 8; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LAAAEKAG 146
DB 59 LAAAEKAG 66

RESULT 5
Q94959 PRELIMINARY; PRT; 364 AA.
AC Q94959;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE KIAA0883 PROTEIN.
GN KIAA0883.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
```

DR EMBL: AB020690; BAA74906.1; -  
SQ SEQUENCE 364 AA; 41509 MW; 6E417AD96E3F0E93 CRC64;

Query Match 3.5%; Score 8; DB 4; Length 364;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ELLAHLG 211  
Db 140 ELLAHLG 147  
|||||

RESULT 6  
Q9IBF0 PRELIMINARY; PRT; 392 AA.  
AC Q9IBF0  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE WILMS' TUMOR PROTEIN.  
GN EWT1.  
OS Anguilla japonica (Japanese eel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguilloidei;  
OC Anguillidae; Anguilla.  
OX NCBI\_TaxID=7937;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Nakatsuru Y., Minami K., Yoshikawa A., Zhu J.J., Oda H., Masahito P.,  
RA Okamoto N., Nakamura Y., Ishikawa T.,  
RT "Eel Wt1 sequence and expression in spontaneous nephroblastomas in  
Japanese eel."  
RL Gene 245:245-251(2000).  
DR EMBL: AB030741; BAA90558.1; -  
SQ SEQUENCE 392 AA; 43892 MW; DAE1A84828F43DF6 CRC64;

Query Match 3.5%; Score 8; DB 13; Length 392;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ACRYGAFG 182  
Db 92 ACRYGAFG 99  
|||||

RESULT 7  
Q9PUT7 PRELIMINARY; PRT; 419 AA.  
AC Q9PUT7  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE WILMS' TUMOR SUPPRESSOR.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith S.I., Down M., Power M., Boyd A.W.;  
RT "Isolation and characterization of a cDNA encoding zebrafish (Danio  
rerio) WT-1."  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF144550; AAF00123.1; -  
DR HSP: P08046; 1AA  
DR INTERPRO: IPR000822; -  
DR PFAM: PF00096; -  
DR PFAM: PF00096; zf-C2H2; 4.  
DR PFAM: PF02165; WT1; 3.

DR PRINTS: PR00049; WILMSTUMOUR.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 4.  
SQ SEQUENCE 419 AA; 46925 MW; 606ADF6A619EED CRC64;

Query Match 3.5%; Score 8; DB 13; Length 419;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ACRYGAFG 182  
Db 91 ACRYGAFG 98  
|||||

RESULT 8  
Q9W611 PRELIMINARY; PRT; 426 AA.  
AC Q9W611  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE WT1.  
OS Cynops pyrrhogaster (Japanese common newt).  
OC Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae; Cynops.  
OX NCBI\_TaxID=8330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Nakayama Y., Yamamoto T., Matsuda Y., Abe S.I.;  
RT "Cloning of cDNA for newt Wt1 and the differential expression during  
spermatogenesis of the Japanese newt, Cynops pyrrhogaster."  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB013888; BAA76399.1; -  
DR HSP: P08046; 1AA  
DR INTERPRO: IPR000822; -  
DR PFAM: PF00096; zf-C2H2; 4.  
DR PRINTS: PR00049; WILMSTUMOUR.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 4.  
DR ZINC-FINGER: Metal-binding; DNA-binding.  
KW ZINC-FINGER; Metal-binding; DNA-binding.  
SQ SEQUENCE 426 AA; 47590 MW; 46A0B158A9F79C6C CRC64;

Query Match 3.5%; Score 8; DB 13; Length 426;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ACRYGAFG 182  
Db 90 ACRYGAFG 97  
|||||

RESULT 9  
O85409 PRELIMINARY; PRT; 83 AA.  
ID O85409  
AC O85409  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
GN OPR1.  
OS Pseudomonas mendocina.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LMG 1223;  
RX MEDLINE=89313294; PubMed=2473376;  
RA Cornellis P., Boula A., Belarbi A., Guyonvarch A., Kammerer B.,  
RA Hannaert V., Hubert J.C.;  
RT "Cloning and analysis of the gene for the major outer membrane

RT lipoprotein from Pseudomonas aeruginosa.;  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 1223;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065948; AAD03503.1; -;  
 KW Lipoprotein.  
 FT NON\_TER  
 SQ SEQUENCE 83 AA; 8843 MW; EF0F54D80801DBA1 CRC64;  
 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 113 MLEKASR 119  
 Db 76 MLEKASR 82  
 RESULT 10  
 O85410  
 ID O85410 PRELIMINARY; PRT; 83 AA.  
 AC O85410;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas oleovorans.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=301;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2229;  
 RA Cornelis P., Bouia A., Belarbi A., Guyonvarch A., Kammerer B.,  
 RA Hannaert V., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2229;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065949; AAD03504.1; -;  
 KW Lipoprotein.  
 FT NON\_TER  
 SQ SEQUENCE 83 AA; 8801 MW; EF0123A10676A2A1 CRC64;  
 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 113 MLEKASR 119  
 Db 76 MLEKASR 82  
 RESULT 11  
 O85411  
 ID O85411 PRELIMINARY; PRT; 83 AA.

AC O85411;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas pseudocaligenes.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=330;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 6036;  
 RX MEDLINE=89313294; PubMed=2473376;  
 RA Cornelis P., Bouia A., Belarbi A., Guyonvarch A., Kammerer B.,  
 RA Hannaert V., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 6036;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065950; AAD03505.1; -;  
 KW Lipoprotein.  
 FT NON\_TER  
 SQ SEQUENCE 83 AA; 8784 MW; EF0F54DA8201DBA1 CRC64;  
 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 113 MLEKASR 119  
 Db 76 MLEKASR 82  
 RESULT 12  
 O85412  
 ID O85412 PRELIMINARY; PRT; 83 AA.  
 AC O85412;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas alcaligenes.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=43263;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 1224;  
 RX MEDLINE=89313294; PubMed=2473376;  
 RA Hannaert V., Bouia A., Belarbi A., Guyonvarch A., Kammerer B.,  
 RA Cornelis P., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 1224;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065951; AAD03506.1; -;

KW Lipoprotein. 83 83  
 FT NON\_TER 83 AA; 8784 MW; EF0F54DA8201DBA1 CRC64;  
 SQ SEQUENCE 83 AA; 8784 MW; EF0F54DA8201DBA1 CRC64;  
 Query Match 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
 DB 76 MLEKASR 82

RESULT 13  
 ID 085413 PRELIMINARY; PRT; 83 AA.  
 AC 085413; 1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas pseudoalcaligenes.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=330;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 5516;  
 RA De Vos D., Bouton C., Sarniguet A., Guyonvarch A., Kammerer B.,  
 RA Cornelis P., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa.";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 5516;  
 RA De Vos D., Bouton C., Sarniguet A., Guyonvarch A., Kammerer B.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065952; AAD03507.1;  
 KW Lipoprotein. 83 83  
 FT NON\_TER 83 AA; 8829 MW; E80854DF0F01DBA1 CRC64;  
 SQ SEQUENCE 83 AA; 8829 MW; E80854DF0F01DBA1 CRC64;  
 Query Match 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
 DB 76 MLEKASR 82

RESULT 14  
 ID 085414 PRELIMINARY; PRT; 83 AA.  
 AC 085414; 1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 5516;  
 RA De Vos D., Bouton C., Sarniguet A., Guyonvarch A., Kammerer B.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065952; AAD03507.1;  
 KW Lipoprotein. 83 83  
 FT NON\_TER 83 AA; 8829 MW; E80854DF0F01DBA1 CRC64;  
 SQ SEQUENCE 83 AA; 8829 MW; E80854DF0F01DBA1 CRC64;  
 Query Match 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
 DB 76 MLEKASR 82

RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2257;  
 RX MEDLINE=89313294; PubMed=2473376;  
 RA Cornelis P., Bouia A., Belarbi A., Guyonvarch A., Kammerer B.,  
 RA Hannaert V., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa.";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2257;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065953; AAD03508.1;  
 KW Lipoprotein. 83 83  
 FT NON\_TER 83 AA; 8823 MW; FB5F54D8131BDEBE CRC64;  
 SQ SEQUENCE 83 AA; 8823 MW; FB5F54D8131BDEBE CRC64;

Query Match 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
 DB 76 MLEKASR 82

RESULT 15  
 ID 085415 PRELIMINARY; PRT; 83 AA.  
 AC 085415; 1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE MAJOR OUTER MEMBRANE LIPOPROTEIN I (FRAGMENT).  
 GN OPRI.  
 OS Pseudomonas cichorii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=36746;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2162;  
 RX MEDLINE=89313294; PubMed=2473376;  
 RA Cornelis P., Bouia A., Belarbi A., Guyonvarch A., Kammerer B.,  
 RA Hannaert V., Hubert J.C.;  
 RT "Cloning and analysis of the gene for the major outer membrane  
 RT lipoprotein from Pseudomonas aeruginosa.";  
 RL Mol. Microbiol. 3:421-428(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LMG 2162;  
 RA De Vos D., Bouton C., Sarniguet A., De Vos P., Vauterin M.,  
 RA Cornelis P.;  
 RT "Sequence diversity of the oprI gene coding for the major outer  
 RT membrane lipoprotein I among rRNA group I pseudomonads.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065954; AAD03509.1;  
 KW Lipoprotein. 83 83  
 FT NON\_TER 83 AA; 8827 MW; FB5F436EC860DBBA CRC64;  
 SQ SEQUENCE 83 AA; 8827 MW; FB5F436EC860DBBA CRC64;

Query Match 3.1%; Score 7; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 MLEKASR 119  
 DB 76 MLEKASR 82

Fri Apr 20 10:52:41 2001

us-09-441-723-1.rspt

Page 6

Db 76 MLEKASR 82

Search completed: April 19, 2001, 13:00:36  
Job time: 119 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:58:12 ; Search time 4.15 Seconds  
(without alignments)  
137.860 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGPLPRTVELFDVLSYWS.....AHLGKWMGPIPPAVNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 12259 seqs, 2531507 residues

Word size : 0

Total number of hits satisfying chosen parameters: 12259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.1	261	5	US-09-739-449-11148
2	7	3.1	688	6	US-60-248-505-865
3	6	2.7	62	5	US-09-739-449-10531
4	6	2.7	161	5	US-09-739-449-12044
5	6	2.7	207	5	US-09-739-449-9529
6	6	2.7	215	5	US-09-739-449-9343
7	6	2.7	235	5	US-09-739-449-9873
8	6	2.7	235	5	US-09-739-449-10784
9	6	2.7	245	5	US-09-739-449-10257
10	6	2.7	258	5	US-09-739-449-10586
11	6	2.7	305	5	US-09-739-449-9486
12	6	2.7	335	5	US-09-739-449-11093
13	6	2.7	325	5	US-09-739-449-12418
14	6	2.7	327	5	US-09-739-449-13257
15	6	2.7	329	5	US-09-739-449-12591
16	6	2.7	337	5	US-09-739-449-11560
17	6	2.7	338	5	US-09-739-449-12951
18	6	2.7	447	5	US-09-739-449-11196
19	6	2.7	456	5	US-09-739-449-13245
20	6	2.7	474	5	US-09-739-449-11298
21	6	2.7	496	6	US-60-248-505-1273
22	6	2.7	570	6	US-60-248-505-980
23	6	2.7	730	6	US-60-248-505-921
24	6	2.7	926	5	US-09-421-124-187
25	6	2.7	926	5	US-09-421-124-187
26	6	2.7	1116	5	US-09-739-449-11359
27	6	2.7	2882	6	US-60-248-505-1095

28	6	2.7	3064	6	US-60-248-505-1172	Sequence 1172, Ap
29	6	2.7	4044	6	US-60-248-505-791	Sequence 791, App
30	5	2.2	10	5	US-09-819-308-40	Sequence 40, Appl
31	5	2.2	18	5	US-09-734-520-4	Sequence 4, Appl
32	5	2.2	20	5	US-09-812-528-18	Sequence 18, Appl
33	5	2.2	21	5	US-09-234-832-23	Sequence 23, Appl
34	5	2.2	35	5	US-09-315-304B-710	Sequence 710, App
35	5	2.2	36	5	US-09-315-304B-89	Sequence 89, Appl
36	5	2.2	39	5	US-09-809-391-598	Sequence 598, App
37	5	2.2	39	5	US-09-315-304B-1456	Sequence 1456, Ap
38	5	2.2	39	5	US-09-315-304B-1530	Sequence 1530, Ap
39	5	2.2	42	5	US-09-601-600-180	Sequence 180, App
40	5	2.2	45	5	US-09-739-449-11782	Sequence 11782, A
41	5	2.2	47	5	US-09-739-449-12291	Sequence 12291, A
42	5	2.2	50	5	US-09-739-449-8955	Sequence 8955, Ap
43	5	2.2	52	5	US-09-739-449-12800	Sequence 12800, A
44	5	2.2	76	5	US-09-739-449-9509	Sequence 9509, Ap
45	5	2.2	78	5	US-09-739-449-12817	Sequence 12817, A

## ALIGNMENTS

RESULT 1  
US-09-739-449-11148  
; Sequence 11148, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 11148  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-11148

Query Match 3.1%; Score 7; DB 5; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAEEKAG 146  
Db 50 AAEEKAG 56

RESULT 2  
US-60-248-505-865  
; Sequence 865, Application US/60248505  
; GENERAL INFORMATION:  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; FILE REFERENCE: G1000918  
; CURRENT APPLICATION NUMBER: US/60/248,505  
; CURRENT FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 1998  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 865  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Human  
US-60-248-505-865

Query Match 3.1%; Score 7; DB 6; Length 688;

Best Local Similarity 100.0%; Pred. No. 6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 LQLRSL 43  
      |||||  
Db 50 LQLRSL 56

## RESULT 3

US-09-739-449-10531  
; Sequence 10531, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 10531  
; LENGTH: 62  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-10531

Query Match 2.7%; Score 6; DB 5; Length 62;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 QLKETT 172  
      |||||  
Db 5 QLKETT 10

## RESULT 4

US-09-739-449-12044  
; Sequence 12044, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 12044  
; LENGTH: 161  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-12044

Query Match 2.7%; Score 6; DB 5; Length 161;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TAVNLE 109  
      |||||  
Db 153 TAVNLE 158

## RESULT 5

US-09-739-449-9529  
; Sequence 9529, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 9529  
; LENGTH: 207  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-9529

Query Match 2.7%; Score 6; DB 5; Length 207;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 ILAAAE 143  
      |||||  
Db 15 ILAAAE 20

## RESULT 6

US-09-739-449-9343  
; Sequence 9343, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 9343  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-9343

Query Match 2.7%; Score 6; DB 5; Length 215;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LAAAEK 144  
      |||||  
Db 159 LAAAEK 164

## RESULT 7

US-09-739-449-9873  
; Sequence 9873, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 9873  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-9873

Query Match 2.7%; Score 6; DB 5; Length 235;  
Best Local Similarity 100.0%; Pred. No. 23;



Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 LSAMRF 102  
|||||  
Db 143 LSAMRF 148

## RESULT 8

US-09-739-449-10784  
; Sequence 10784, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 10784  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-10784

Query Match 2.7%; Score 6; DB 5; Length 235;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 SLSAMR 101  
|||||  
Db 57 SLSAMR 62

## RESULT 9

US-09-739-449-10257  
; Sequence 10257, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 10257  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-10257

Query Match 2.7%; Score 6; DB 5; Length 245;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 QSLAA 141  
|||||  
Db 182 QSLAA 187

## RESULT 10

US-09-739-449-10586  
; Sequence 10586, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C

Query Match 2.7%; Score 6; DB 5; Length 315;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 10586  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-10586

Query Match 2.7%; Score 6; DB 5; Length 258;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 MLFGSD 201  
|||||  
Db 216 MLFGSD 221

## RESULT 11

US-09-739-449-9486  
; Sequence 9486, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 9486  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-9486

Query Match 2.7%; Score 6; DB 5; Length 305;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 LSVMLE 93  
|||||  
Db 219 LSVMLE 224

## RESULT 12

US-09-739-449-11093  
; Sequence 11093, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 11093  
; LENGTH: 315  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-11093

Query Match 2.7%; Score 6; DB 5; Length 315;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 GAGLP 184  
 Db 250 GAGLP 255

## RESULT 13

US-09-739-449-12418  
 ; Sequence 12418, Application US/09739449  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15490)C  
 ; CURRENT APPLICATION NUMBER: US/09/739,449  
 ; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 09/514,000  
 ; PRIOR FILING DATE: 2000-02-23  
 ; NUMBER OF SEQ ID NOS: 13351  
 ; SEQ ID NO 12418  
 ; LENGTH: 325  
 ; TYPE: PRT  
 ; ORGANISM: Agrobacterium tumefaciens  
 US-09-739-449-12418

Query Match 2.7%; Score 6; DB 5; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAEKA 145  
 Db 81 AAAEKA 86

## RESULT 14

US-09-739-449-13257  
 ; Sequence 13257, Application US/09739449  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15490)C  
 ; CURRENT APPLICATION NUMBER: US/09/739,449  
 ; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 09/514,000  
 ; PRIOR FILING DATE: 2000-02-23  
 ; NUMBER OF SEQ ID NOS: 13351  
 ; SEQ ID NO 13257  
 ; LENGTH: 327  
 ; TYPE: PRT  
 ; ORGANISM: Agrobacterium tumefaciens  
 US-09-739-449-13257

Query Match 2.7%; Score 6; DB 5; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 QGILEK 158  
 Db 59 QGILEK 64

## RESULT 15

US-09-739-449-12591  
 ; Sequence 12591, Application US/09739449  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15490)C  
 ; CURRENT APPLICATION NUMBER: US/09/739,449

; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 09/514,000  
 ; PRIOR FILING DATE: 2000-02-23  
 ; NUMBER OF SEQ ID NOS: 13351  
 ; SEQ ID NO 12591  
 ; LENGTH: 329  
 ; TYPE: PRT  
 ; ORGANISM: Agrobacterium tumefaciens  
 US-09-739-449-12591

Query Match 2.7%; Score 6; DB 5; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAEEKAG 146  
 Db 98 AAEEKAG 103

Search completed: April 19, 2001, 13:00:08  
 Job time: 116 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2001, 12:58:02 : Search time 44.47 Seconds  
(without alignments)  
817.475 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 226  
Sequence: 1 MGLPLRTVELFYDLSPYSW .....AHLLEKWMGPIPPAVNARL 226

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1009251 seqs, 160854530 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1009251

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	226	100.0	226	18	US-09-441-723-1
2	150	66.4	150	23	US-60-147-499-1666
3	97	42.9	97	1	PCT-US00-06824-176
4	92	40.7	120	23	US-60-245-201-153
5	92	40.7	120	23	US-60-245-201-226
6	16	7.1	226	18	US-09-441-723-3
7	12	5.3	97	1	PCT-US00-06824-175
8	8	3.5	91	16	US-09-252-991A-23967
9	8	3.5	173	18	US-09-489-039A-7934
10	8	3.5	179	1	PCT-US00-05918-529

11	8	3.5	195	15	US-09-189-527-7	Sequence 7, Appli
12	8	3.5	195	15	US-09-189-527-7	Sequence 7, Appli
13	8	3.5	814	6	US-08-233-642A-40	Sequence 40, Appli
14	7	3.1	46	23	US-60-186-662-1013	Sequence 1013, Ap
15	7	3.1	55	8	US-08-424-550A-493	Sequence 493, App
16	7	3.1	55	8	US-08-424-550A-493	Sequence 493, App
17	7	3.1	55	8	US-08-469-260-493	Sequence 493, App
18	7	3.1	55	8	US-08-469-260A-493	Sequence 493, App
19	7	3.1	55	8	US-08-486-749-493	Sequence 493, App
20	7	3.1	55	8	US-08-488-445A-493	Sequence 493, App
21	7	3.1	55	8	US-08-488-446-493	Sequence 493, App
22	7	3.1	63	9	US-08-572-447B-9	Sequence 9, Appli
23	7	3.1	63	16	US-09-267-747-9	Sequence 24513, A
24	7	3.1	75	18	US-09-417-507-24513	Sequence 22114, A
25	7	3.1	78	16	US-09-248-796-22114	Sequence 3205, Ap
26	7	3.1	87	23	US-60-196-712-3205	Sequence 2061, Ap
27	7	3.1	91	23	US-60-196-711-2061	Sequence 2062, Ap
28	7	3.1	93	23	US-60-196-711-2062	Sequence 26172, A
29	7	3.1	104	16	US-09-252-991A-26172	Sequence 854, App
30	7	3.1	130	1	PCT-US00-05918-854	Sequence 36291, A
31	7	3.1	132	16	US-09-270-767-36291	Sequence 51508, A
32	7	3.1	133	4	US-08-047-167-3	Sequence 3, Appli
33	7	3.1	133	12	US-08-802-741-19	Sequence 19, Appli
34	7	3.1	136	16	US-09-270-767-37986	Sequence 37986, A
35	7	3.1	141	1	PCT-US99-22853B-972	Sequence 971, App
36	7	3.1	142	1	PCT-US99-22853B-971	Sequence 340, App
37	7	3.1	161	23	US-60-207-583-340	Sequence 10, Appli
38	7	3.1	204	10	US-08-691-814-10	Sequence 10, Appli
39	7	3.1	204	10	US-08-691-814A-10	Sequence 4, Appli
40	7	3.1	204	16	US-09-250-609-4	Sequence 9, Appli
41	7	3.1	204	16	US-09-250-609-9	Sequence 4, Appli
42	7	3.1	204	16	US-09-250-609-4	Sequence 9, Appli
43	7	3.1	204	16	US-09-250-609-9	Sequence 4, Appli
44	7	3.1	204	16	US-09-250-611-4	Sequence 4, Appli
45	7	3.1	204	16	US-09-250-611-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-441-723-1  
Sequence 1, Application US/09441723  
GENERAL INFORMATION:  
APPLICANT: Shah, Purvi  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/441,723  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/978,174  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0430 US

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-855-0555  
 TELEFAX: 650-845-4166  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 226 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: BLADTUT04  
 CLONE: 1554593  
 US-09-441-723-1

Query Match 100.0%; Score 226; DB 18; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-224;  
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPLPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 DB 1 MGPLPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 QY 61 RKGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
 DB 61 RKGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
 QY 121 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGA 180  
 DB 121 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGA 180  
 QY 181 FGLPTITVAHVQDQTHMFGSDRMELLAHLGKWMGPIPPAVNARL 226  
 DB 181 FGLPTITVAHVQDQTHMFGSDRMELLAHLGKWMGPIPPAVNARL 226

RESULT 2  
 US-60-147-499-4666  
 ; Sequence 4666, Application US/60147499  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSER:054PR2  
 ; CURRENT APPLICATION NUMBER: US/60/147,499  
 ; CURRENT FILING DATE: 1999-08-05  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 4666  
 ; LENGTH: 150  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-60-147-499-4666

Query Match 66.4%; Score 150; DB 23; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-146;  
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPLPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 DB 1 MGPLPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 QY 61 RKGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
 DB 61 RKGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
 QY 121 LWMRVSRNEDITEPQSILAAAEKAGMSAE 150  
 DB 121 LWMRVSRNEDITEPQSILAAAEKAGMSAE 150

RESULT 3  
 PCT-US00-06824-176  
 ; Sequence 176, Application PC/TUS0006824  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc.  
 ; TITLE OF INVENTION: 47 Human Secreted Proteins  
 ; FILE REFERENCE: PS511PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US00/06824  
 ; CURRENT FILING DATE: 2000-03-16  
 ; EARLIER APPLICATION NUMBER: 60/125,359  
 ; EARLIER FILING DATE: 1999-03-19  
 ; EARLIER APPLICATION NUMBER: 60/168,664  
 ; EARLIER FILING DATE: 1999-12-03  
 ; NUMBER OF SEQ ID NOS: 180  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 176  
 ; LENGTH: 97  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 PCT-US00-06824-176

Query Match 42.9%; Score 97; DB 1; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-91;  
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 130 EDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGAFGLPITVAH 189  
 DB 1 EDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGAFGLPITVAH 60  
 QY 190 VDGQTHMFGSDRMELLAHLGKWMGPIPPAVNARL 226  
 DB 61 VDGQTHMFGSDRMELLAHLGKWMGPIPPAVNARL 97

RESULT 4  
 US-60-245-201-153  
 ; Sequence 153, Application US/60245201  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Beasley, Ellen  
 ; TITLE OF INVENTION: ISOLATED HUMAN PHASE II  
 ; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, NUCLEIC ACID MOLECULES ENCODING  
 ; FILE REFERENCE: CL000879  
 ; CURRENT APPLICATION NUMBER: US/60/245,201  
 ; CURRENT FILING DATE: 2000-11-03  
 ; NUMBER OF SEQ ID NOS: 381  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 153  
 ; LENGTH: 120  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 US-60-245-201-153

Query Match 40.7%; Score 92; DB 23; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-86;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 135 PQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGAFGLPITVAHVQGT 194  
 DB 29 PQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLEKTEAACRYGAFGLPITVAHVQGT 88  
 QY 195 HMLFGSDRMELLAHLGKWMGPIPPAVNARL 226  
 DB 89 HMLFGSDRMELLAHLGKWMGPIPPAVNARL 120

RESULT 5  
 US-60-245-201-226  
 ; Sequence 226, Application US/60245201  
 ; GENERAL INFORMATION:



Db 79 AEQAQGLL 86

## RESULT 9

US-09-489-039A-7934  
; Sequence 734, Application US/09489039A  
; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: 2709.2004001

; PRIOR FILING DATE: 2000-01-27 US/09/489,039A

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7934

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7934

## Query Match

Best Local Similarity 3.5%; Score 8; DB 18; Length 173;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 LAAAEKAG 146

|||||||

Db 120 LAAAEKAG 127

## RESULT 10

PCT-US00-05918-529

; Sequence 529, Application PC/TUS0005918

; GENERAL INFORMATION:

; APPLICANT: Craig Rosen,

; APPLICANT: Steve Ruben

; TITLE OF INVENTION: Human Lung Cancer Associated Gene Sequences and Polypeptides

; FILE REFERENCE: PA104PCT

; CURRENT APPLICATION NUMBER: PCT/US00/05918

; EARLIER APPLICATION NUMBER: 2000-03-08

; EARLIER FILING DATE: 2000-03-12

; NUMBER OF SEQ ID NOS: 896

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 529

; LENGTH: 179

; TYPE: PRT

; ORGANISM: Homo sapiens

PCT-US00-05918-529

## Query Match

Best Local Similarity 3.5%; Score 8; DB 1; Length 179;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 SLSAMREL 103

|||||||

Db 25 SLSAMREL 32

## RESULT 11

US-09-189-527-7

; Sequence 7, Application US/09189527

; GENERAL INFORMATION:

; APPLICANT: Jerome B. Posner

; APPLICANT: Josep O. Dalmau

; TITLE OF INVENTION: Ma Family Polypeptides and Anti-Ma Antibodies

; FILE REFERENCE: SLK98-01

; CURRENT APPLICATION NUMBER: US/09/189,527

; CURRENT FILING DATE: 1998-11-10

; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 7

; LENGTH: 195

; TYPE: PRT

; ORGANISM: homo sapiens

US-09-189-527-7

## Query Match

Best Local Similarity 3.5%; Score 8; DB 15; Length 195;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ELLAHLGL 211

|||||||

Db 139 ELLAHLGL 146

## RESULT 12

US-09-189-527-7

; Sequence 7, Application US/09189527A

; GENERAL INFORMATION:

; APPLICANT: Jerome B. Posner

; APPLICANT: Josep O. Dalmau

; TITLE OF INVENTION: Ma Family Polypeptides and Anti-Ma

; FILE REFERENCE: SLK98-01

; CURRENT APPLICATION NUMBER: US/09/189,527A

; CURRENT FILING DATE: 1998-11-10

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 7

; LENGTH: 195

; TYPE: PRT

; ORGANISM: homo sapiens

US-09-189-527-7

## Query Match

Best Local Similarity 3.5%; Score 8; DB 15; Length 195;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ELLAHLGL 211

|||||||

Db 139 ELLAHLGL 146

## RESULT 13

US-08-233-642A-40

; Sequence 40, Application US/08233642A

; GENERAL INFORMATION:

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Clouthier, Sharon C.

; APPLICANT: Doran, James L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; STREET: Seed and Berry

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/08/233,642A  
FILING DATE: 26-APR-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Joshua  
REGISTRATION NUMBER: 35,570  
REFERENCE/DOCKET NUMBER: 920043.403C3  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEEDANBERRY  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 814 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-233-642A-40

Query Match 3.5%; Score 8; DB 6; Length 814;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 NOLKETTE 173  
Db 60 NOLKETTE 67

RESULT 14  
US-60-186-662-1013  
Sequence 1013, Application US/60186662  
GENERAL INFORMATION:  
APPLICANT: Bonazzi, Vivien  
TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: CL000314  
CURRENT APPLICATION NUMBER: US/60/186,662  
CURRENT FILING DATE: 2000-03-03  
NUMBER OF SEQ ID NOS: 1080  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1013  
LENGTH: 46  
TYPE: PRT  
ORGANISM: HUMAN  
US-60-186-662-1013

Query Match 3.1%; Score 7; DB 23; Length 46;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 PGLLPR 61  
Db 40 PGLLPR 46

RESULT 15  
US-08-424-550A-493  
Sequence 493, Application US/08424550A  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMI J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUERHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS

NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: POREBSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 493:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 55 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-424-550A-493

Query Match 3.1%; Score 7; DB 8; Length 55;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 GSLSAMR 101  
Db 42 GSLSAMR 48

Search completed: April 19, 2001, 13:00:00  
Job time: 118 sec

